

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 22:30:21 ; Search time 973.268 Seconds
(without alignments)
8940.745 Million cell updates/sec

Title: US-10-029-359A-1

Perfect score: 299

Sequence: 1 gcaagaggagcagcgccg.....tcgctgaaacacctctattag 299

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg_hum.*

31: em.htg_inv.*

32: em.htg_other.*

33: em.htg_mus.*

34: em.htg_pln.*

35: em.htg_rod.*

36: em.htg_mam.*

37: em.htg_vrt.*

38: em_sy.*

39: em_higo_hum.*

40: em_higo_mus.*

41: em_higo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|--------------------|
| C 1 | 283.8 | 94.9 | 156337 | 2 | AC009938 | AC009938 Homo sapi |
| C 2 | 283.8 | 94.9 | 169898 | 2 | AC016585 | AC016585 Homo sapi |
| C 3 | 283.8 | 94.9 | 172574 | 2 | AC011121 | AC011121 Homo sapi |
| C 4 | 283.8 | 94.9 | 182823 | 9 | AC016638 | AC016638 Homo sapi |
| C 5 | 283.8 | 94.9 | 217807 | 2 | AC079469 | AC079469 Homo sapi |
| C 6 | 283.8 | 94.9 | 217807 | 2 | AC079469 | AC079469 Homo sapi |
| C 7 | 236.4 | 79.1 | 1794 | 6 | AF175160 | AF175160 Sequence |
| C 8 | 231 | 77.3 | 930 | 9 | AF155582 | AF155582 Homo sapi |
| C 9 | 231 | 77.3 | 1092 | 6 | HS132443 | HS132443 Homo sapi |
| C 10 | 231 | 77.3 | 1092 | 6 | AX175170 | AX175170 Sequence |
| C 11 | 231 | 77.3 | 1092 | 9 | HS243256 | HS243256 Homo sapi |
| C 12 | 206.6 | 69.1 | 115954 | 9 | HS278960 | HS278960 Homo sapi |
| C 13 | 194 | 64.9 | 1438 | 10 | AC005532 | AC005532 Homo sapi |
| C 14 | 194 | 64.9 | 1469 | 6 | AF157962 | AF157962 Mus muscu |
| C 15 | 192.6 | 64.4 | 1092 | 6 | AX175164 | AX175164 Sequence |
| C 16 | 190.4 | 63.7 | 123070 | 9 | AX175172 | AX175172 Sequence |
| C 17 | 190.4 | 63.7 | 307755 | 2 | AC020656 | AC020656 Homo sapi |
| C 18 | 189.2 | 63.3 | 1405 | 10 | AC127894 | AC127894 Homo sapi |
| C 19 | 189.2 | 63.3 | 1440 | 6 | AF157963 | AF157963 Rattus no |
| C 20 | 188.8 | 63.1 | 170227 | 2 | AX175162 | AX175162 Sequence |
| C 21 | 187.8 | 62.8 | 1092 | 6 | AC019161 | AC019161 Homo sapi |
| C 22 | 169.8 | 56.8 | 185872 | 2 | AX175171 | AX175171 Sequence |
| C 23 | 163.4 | 54.6 | 194038 | 2 | AC068810 | AC068810 Mus muscu |
| C 24 | 145.8 | 48.8 | 171529 | 2 | AC095494 | AC095494 Rattus no |
| C 25 | 111.4 | 37.3 | 146790 | 2 | AC023021 | AC023021 Homo sapi |
| C 26 | 83.4 | 27.9 | 167996 | 9 | AC118920 | AC118920 Rattus no |
| C 27 | 83.4 | 27.9 | 191655 | 9 | AC021753 | AC021753 Homo sapi |
| C 28 | 78.8 | 26.4 | 153472 | 9 | AC020661 | AC020661 Homo sapi |
| C 29 | 78.8 | 26.4 | 165720 | 9 | AC011179 | AC011179 Homo sapi |
| C 30 | 67.2 | 22.5 | 183416 | 10 | AC025524 | AC025524 Homo sapi |
| C 31 | 67.2 | 22.5 | 229117 | 2 | AL671269 | AL671269 Mouse DNA |
| C 32 | 53.2 | 17.8 | 1708 | 10 | AL807383 | AL807383 Mus muscu |
| C 33 | 50.8 | 17.0 | 121200 | 2 | BC025899 | BC025899 Mus muscu |
| C 34 | 42 | 14.0 | 213308 | 2 | AC112336 | AC112336 Rattus no |
| C 35 | 42 | 14.0 | 217953 | 2 | AC079164 | AC079164 Mus muscu |
| C 36 | 41.6 | 13.9 | 173665 | 2 | AC122334 | AC122334 Mus muscu |
| C 37 | 40.8 | 13.6 | 187431 | 9 | AC115770 | AC115770 Mus muscu |
| C 38 | 40.4 | 13.5 | 68573 | 2 | AC107219 | AC107219 Homo sapi |
| C 39 | 40.2 | 13.4 | 176113 | 2 | AC107791 | AC107791 Mus muscu |
| C 40 | 40.2 | 13.4 | 191821 | 2 | AC103226 | AC103226 Rattus no |
| C 41 | 40 | 13.4 | 173430 | 2 | AC123475 | AC123475 Rattus no |
| C 42 | 39.8 | 13.3 | 167274 | 2 | AC112689 | AC112689 Mus muscu |
| C 43 | 39.2 | 13.1 | 3322 | 8 | AC114246 | AC114246 Rattus no |
| C 44 | 39 | 13.0 | 164581 | 2 | R1CFUJI | D50602 Rice mRNA f |
| C 45 | 38.8 | 13.0 | 148950 | 2 | AC126794 | AC126794 Medicago |
| | | | | | AC108583 | Rattus no |

ALIGNMENTS

RESULT 1
AC009938/c AC009938 156337 bp DNA linear HTG 13-JUL-2000
LOCUS Homo sapiens clone RP11-115B14, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC009938
ACCESSION AC009938
VERSION AC009938.3 GI:9113959
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 156337)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-115B14
JOURNAL Unpublished

REFERENCE
AUTHORS

2 (bases 1 to 156337)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
 Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A.,
 Castle, A., Cerny, J., Collangelo, M., Collins, S., Collymore, A.,
 Cooke, P., Deaerillano, K., Depayre, E., Devon, K., Dewar, K.,
 Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C.,
 Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,
 Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L.,
 Karas, A., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P.,
 Marquis, N., McEwan, P., McGuck, A., McKernan, K., McLaughlin, J.,
 Melidrim, J., Molla, M., Morris, W., Morrow, J., Mychaleckyl, J.,
 Naylor, J., Nifloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,
 Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,
 Testaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,
 Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.
 Direct Submission
 Submitted (08-SEP-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 13, 2000 this sequence version replaced gi:6978143.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L1978
 Center clone name: L15_B_14

* NOTE: This record contains 149 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 995: contig of 995 bp in length
 * 996 1095: gap of 100 bp
 * 1096 2085: contig of 970 bp in length
 * 2086 2165: gap of 100 bp
 * 2166 3117: contig of 952 bp in length
 * 3118 3217: gap of 100 bp
 * 3218 4168: contig of 951 bp in length
 * 4169 4268: gap of 100 bp
 * 4269 5236: contig of 968 bp in length
 * 5237 5336: gap of 100 bp
 * 5337 6309: contig of 973 bp in length
 * 6310 6409: gap of 100 bp
 * 6410 7365: contig of 956 bp in length
 * 7366 7465: gap of 100 bp
 * 7466 8417: contig of 952 bp in length
 * 8418 8517: gap of 100 bp
 * 8518 9471: contig of 954 bp in length
 * 9472 9571: gap of 100 bp
 * 9572 10511: contig of 940 bp in length
 * 10512 10611: gap of 100 bp
 * 10612 11556: contig of 945 bp in length
 * 11557 11656: gap of 100 bp
 * 11657 12627: contig of 971 bp in length
 * 12628 12727: gap of 100 bp
 * 12728 13655: contig of 928 bp in length
 * 13656 13755: gap of 100 bp
 * 13756 14734: contig of 979 bp in length
 * 14735 14834: gap of 100 bp
 * 14835 15782: contig of 948 bp in length
 * 15783 15882: gap of 100 bp

15883 16812: contig of 930 bp in length
 * 16813 16912: gap of 100 bp
 * 16913 17855: contig of 943 bp in length
 * 17856 17955: gap of 100 bp
 * 17956 18939: contig of 984 bp in length
 * 18940 19039: gap of 100 bp
 * 19040 19995: contig of 956 bp in length
 * 19996 20095: gap of 100 bp
 * 20096 21022: contig of 927 bp in length
 * 21023 21122: gap of 100 bp
 * 21123 22053: contig of 931 bp in length
 * 22054 22153: gap of 100 bp
 * 22154 23146: contig of 993 bp in length
 * 23147 23246: gap of 100 bp
 * 23247 24206: contig of 960 bp in length
 * 24207 24306: gap of 100 bp
 * 24307 25275: contig of 969 bp in length
 * 25276 25375: gap of 100 bp
 * 25376 26338: contig of 963 bp in length
 * 26339 26438: gap of 100 bp
 * 26439 27420: contig of 982 bp in length
 * 27421 27520: gap of 100 bp
 * 27521 28483: contig of 963 bp in length
 * 28484 28583: gap of 100 bp
 * 28584 29532: contig of 949 bp in length
 * 29533 29632: gap of 100 bp
 * 29633 30571: contig of 939 bp in length
 * 30572 30671: gap of 100 bp
 * 30672 31631: contig of 960 bp in length
 * 31632 31731: gap of 100 bp
 * 31732 32674: contig of 943 bp in length
 * 32675 32774: gap of 100 bp
 * 32775 33735: contig of 961 bp in length
 * 33736 33835: gap of 100 bp
 * 33836 34756: contig of 921 bp in length
 * 34757 34856: gap of 100 bp
 * 34857 35784: contig of 928 bp in length
 * 35785 35884: gap of 100 bp
 * 35885 36818: contig of 934 bp in length
 * 36819 36918: gap of 100 bp
 * 36919 37873: contig of 955 bp in length
 * 37874 37973: gap of 100 bp
 * 37974 38939: contig of 966 bp in length
 * 38940 39039: gap of 100 bp
 * 39040 39953: contig of 914 bp in length
 * 39954 40053: gap of 100 bp
 * 40054 40999: contig of 946 bp in length
 * 41000 41099: gap of 100 bp
 * 41100 42004: contig of 905 bp in length
 * 42005 42104: gap of 100 bp
 * 42105 43018: contig of 914 bp in length
 * 43019 43118: gap of 100 bp
 * 43119 44057: contig of 939 bp in length
 * 44058 44157: gap of 100 bp
 * 44158 45127: contig of 970 bp in length
 * 45128 45227: gap of 100 bp
 * 45228 46166: contig of 939 bp in length
 * 46167 46266: gap of 100 bp
 * 46267 47241: contig of 975 bp in length
 * 47242 47341: gap of 100 bp
 * 47342 48308: contig of 967 bp in length
 * 48309 48408: gap of 100 bp
 * 48409 49391: contig of 983 bp in length
 * 49392 49491: gap of 100 bp
 * 49492 50414: contig of 923 bp in length
 * 50415 50514: gap of 100 bp
 * 50515 51482: contig of 968 bp in length
 * 51483 51582: gap of 100 bp
 * 51583 52526: contig of 944 bp in length
 * 52527 52626: gap of 100 bp
 * 52627 53610: contig of 984 bp in length
 * 53611 53710: gap of 100 bp
 * 53711 54676: contig of 966 bp in length

* 54677 54776: gap of 100 bp
* 54777 55729: contig of 953 bp in length
* 55730 55829: gap of 100 bp
* 55830 56781: contig of 952 bp in length
* 56782 56881: gap of 100 bp
* 56882 57863: contig of 982 bp in length
* 57864 57963: gap of 100 bp
* 57964 58892: contig of 929 bp in length
* 58893 58992: gap of 100 bp
* 58993 59960: contig of 968 bp in length
* 59961 60060: gap of 100 bp
* 60061 61030: contig of 970 bp in length
* 61031 61130: gap of 100 bp
* 61131 62075: contig of 945 bp in length
* 62076 62175: gap of 100 bp
* 62176 63111: contig of 936 bp in length
* 63112 63211: gap of 100 bp
* 63212 64148: contig of 937 bp in length
* 64149 64248: gap of 100 bp
* 64249 65186: contig of 938 bp in length
* 65187 65286: gap of 100 bp
* 65287 66257: contig of 971 bp in length
* 66258 66357: gap of 100 bp
* 66358 67306: contig of 949 bp in length
* 67307 67406: gap of 100 bp
* 67407 68325: contig of 919 bp in length
* 68326 68425: gap of 100 bp
* 68426 69370: contig of 945 bp in length
* 69371 69470: gap of 100 bp
* 69471 70457: contig of 987 bp in length
* 70458 70557: gap of 100 bp
* 70558 71497: contig of 940 bp in length
* 71498 71597: gap of 100 bp
* 71598 72513: contig of 916 bp in length
* 72514 72613: gap of 100 bp
* 72614 73585: contig of 972 bp in length
* 73586 73685: gap of 100 bp
* 73685 74654: contig of 969 bp in length
* 74655 74754: gap of 100 bp
* 74755 75734: contig of 980 bp in length
* 75735 75834: gap of 100 bp

Query Match 94.9%; Score 283.8; DB 2; Length 156337;
Best Local Similarity 99.0%; Pred. No. 1.8e-63;
Matches 296; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 GCAGAGGGAGCCAGCGCGGATGACAGAAATTCACCTTCCGAGATGGCTCTAAATCCT 60
DB 139022 GCAGAGGGAGCCAGCGCGGATGACAG- AAATACACTTCCGAGATGGCTCTAAATCCT 138964
QY 61 GCCTGAATTTTAACTTCCCTCTATGGATCGGCAATAGGGTTTATTTTATTTCTCAGC 120
DB 138963 GCCTGAATTTTAACTTCCCTCTATGGATCGGCAATAGGGTTTATTTTATTTCTCAGC 138904
QY 121 TACTTAGTATTTTGTGGGAGAGAGGGTGCACCCAGACTAATGCTTCTCATATGATC 180
DB 138903 TACTTAGTATTTTGTGGGAGAGAGGGTGCACCCAGACTAATGCTTCTCATATGATC 180
QY 181 CTCATCGAGGATTCAGATGATATGACAGAAATTCATCTAGGAGGACAAATGAATCTCA 240
DB 138843 CTCATCGAGGATTCAGATGATATGACAGAAATTCATCTAGGAGGACAAATGAATCTCA 138784
QY 241 ATGCAGATTCAGCCAACTGAAAGATGAGAAACAGAAATCGCTGAAACCTCTATTAG 299
DB 138783 ATGCAGATTCAGCCAACTGAAAGATGAGAAACAGAAATCGCTGAAACCTCTATTAG 138725

RESULT 2
AC016585/c
LOCUS
DEFINITION Homo sapiens chromosome 5 clone CTD-2601K17, WORKING DRAFT
SEQUENCE, 3 unordered pieces.
AC016585

VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AC016585.6 GI:15290346.
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 169898)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 169898)
DOE Joint Genome Institute.
Direct Submission
Submitted (04-DEC-1999), Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Aug 25, 2001 this sequence version replaced gi:13699586.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 829022
Center clone name: CITB-EL_2601K17

Summary Statistics
Consensus quality: 168838 bases at least Q40
Consensus quality: 169474 bases at least Q30
Consensus quality: 169606 bases at least Q20
Estimated insert size: 167000; pulse field gel estimation
Estimated insert size: 169698; sum-of-contigs estimation
Quality coverage: 7.44 in Q20 bases; pulse field gel estimation
Quality coverage: 7.32 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 10691: contig of 10691 bp in length
* 10692 10791: gap of unknown length
* 10792 53413: contig of 42622 bp in length
* 53414 53513: gap of unknown length
* 53514 169898: contig of 116385 bp in length.
Location/Qualifiers
1. 169898
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2601K17"
/clone_lib="Caltech human BAC library D"
BASE COUNT 49510 a 35344 c 35770 g 49074 t 200 others
ORIGIN

Query Match 94.9%; Score 283.8; DB 2; Length 169898;
Best Local Similarity 99.0%; Pred. No. 1.8e-63;
Matches 296; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 1 GCAGAGGGAGCCAGCGCGGATGACAGAAATTCACCTTCCGAGATGGCTCTAAATCCT 60
DB 112793 GCAGAGGGAGCCAGCGCGGATGACAG- AAATACACTTCCGAGATGGCTCTAAATCCT 112735
QY 61 GCCTGAATTTTAACTTCCCTCTATGGATCGGCAATAGGGTTTATTTTATTTCTCAGC 120
DB 112734 GCCTGAATTTTAACTTCCCTCTATGGATCGGCAATAGGGTTTATTTTATTTCTCAGC 112675
QY 121 TACTTAGTATTTTGTGGGAGAGAGGGTGCACCCAGACTAATGCTTCTCATATGATC 180
DB 112674 TACTTAGTATTTTGTGGGAGAGAGGGTGCACCCAGACTAATGCTTCTCATATGATC 180
QY 181 CTCATCGAGGATTCAGATGATATGACAGAAATTCATCTAGGAGGACAAATGAATCTCA 240

Query Match 94.9%; Score 283.8; DB 2; Length 169898;
Best Local Similarity 99.0%; Pred. No. 1.8e-63;
Matches 296; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 1 GCAGAGGGAGCCAGCGCGGATGACAGAAATTCACCTTCCGAGATGGCTCTAAATCCT 60
DB 112793 GCAGAGGGAGCCAGCGCGGATGACAG- AAATACACTTCCGAGATGGCTCTAAATCCT 112735
QY 61 GCCTGAATTTTAACTTCCCTCTATGGATCGGCAATAGGGTTTATTTTATTTCTCAGC 120
DB 112734 GCCTGAATTTTAACTTCCCTCTATGGATCGGCAATAGGGTTTATTTTATTTCTCAGC 112675
QY 121 TACTTAGTATTTTGTGGGAGAGAGGGTGCACCCAGACTAATGCTTCTCATATGATC 180
DB 112674 TACTTAGTATTTTGTGGGAGAGAGGGTGCACCCAGACTAATGCTTCTCATATGATC 180
QY 181 CTCATCGAGGATTCAGATGATATGACAGAAATTCATCTAGGAGGACAAATGAATCTCA 240

|||||
 Db 112614 CTCATCGAGGCAATTCAGATGATAATGGACAGAAATCATCTAGGAGGACAAATGAATCTCA 112555
 QY 241 ATCAGATTTCTACCAACGTAAGATGAGACACAGAAATCGTGAACCTCTATATAG 299
 Db 112554 ATCAGATTTCTACCAACGTAAGATGAGACACAGATCGTGAACCTCTATATAG 112496

RESULT 3

AC011121/c
 LOCUS AC011121 172574 bp DNA linear HTG 04-SEP-2000
 DEFINITION Homo sapiens clone RP11-364C6, WORKING DRAFT SEQUENCE, 10 unordered
 pieces.

AC011121
 VERSION AC011121.5 GI:9966267
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens

ORGANISM Homo sapiens
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 172574)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS

Unpublished
 2 (bases 1 to 172574)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
 Baldwin, J., Barna, N., Beckerly, R., Boguslavskiy, L., Boukhgalter, B.,
 Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
 Cooke, P., DeArelano, K., Dewar, K., Domino, M., Donegan, L., Doyle, M.,
 Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
 Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heatford, A., Horton, L.,
 Holand, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
 McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
 Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
 Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.

Direct Submission
 Submitted (01-OCT-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Sep 4, 2000 this sequence version replaced gi:7630654.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
 Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information

Center project name: L2174
 Center clone name: 364_C6

----- Summary Statistics
 Sequencing vector: M13; M77815; 96% of reads
 Sequencing vector: Plasmid; n/a; 0.0% of reads

3.84985563041386Chemistry: Dye-terminator Big Dye; 100% of
 reads

Assembly program: Phrap; version 0.960731

Consensus quality: 160604 bases at least Q40

Consensus quality: 165990 bases at least Q30

Consensus quality: 168767 bases at least Q20

Insert size: 182000; agarose-ff

Insert size: 171674; sum-of-ctgigs

Quality coverage: 4.5 in Q20 bases; agarose-ff

Quality cov.

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 10 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence

* as soon as it is available and the accession number will
 * be preserved

1 46314: contig of 46314 bp in length
 46315 46414: gap of 100 bp
 46415 48772: contig of 2358 bp in length
 48773 48872: gap of 100 bp
 48873 51876: contig of 3004 bp in length
 51877 51976: gap of 100 bp
 51977 54631: contig of 2655 bp in length
 54632 54731: gap of 100 bp
 54732 59065: contig of 4334 bp in length
 59066 59165: gap of 100 bp
 59166 63397: contig of 4232 bp in length
 63398 63497: gap of 100 bp
 63498 117527: contig of 54030 bp in length
 117528 117627: gap of 100 bp
 117628 143391: contig of 25764 bp in length
 143392 143491: gap of 100 bp
 143492 167784: contig of 24293 bp in length
 167785 167884: gap of 100 bp
 167885 172574: contig of 4690 bp in length.

FEATURES

source

1..172574
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="RP11-364C6"

/clone_lib="RP11-364C6"
 1..46314

misc_feature

note="assembly_fragment"
 clone_end:SP6
 vector_side:left"

misc_feature

46415..48772
 note="assembly_fragment"
 48873..51876

misc_feature

note="assembly_fragment"
 51977..54631

misc_feature

note="assembly_fragment"
 54732..59065

misc_feature

note="assembly_fragment"
 59166..63397

misc_feature

note="assembly_fragment"
 63498..117527

misc_feature

note="assembly_fragment"
 117628..143391

misc_feature

note="assembly_fragment"
 143492..167784

misc_feature

note="assembly_fragment"
 167885..172574

clone_end:T7
 vector_side:right"

BASE COUNT 50796 a 35214 c 36052 g 49603 t 909 others
 ORIGIN

Query Match 94.9%; Score 283.8; DB 2; Length 172574;

Best Local Similarity 99.0%; Pred. No. 1.8e-63;

Matches 296; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 GCAAGAGGAGGACGCGCGATGACAGAAATTCACATCCGAGATGGCCTCTAAATCCT 60
 Db 80478 GCAAGAGGAGGACGCGCGATGACAG-AAAACACATTTCCGAGATGGCCTCTAAATCCT 80420

QY 61 GCGTGAATTTTAACTTCCTCTATGGATCGGCAATAGGGTTATTTTATTTCTCAGC 120
 Db 80419 GCGTGAATTTTAACTTCCTCTATGGATCGGCAATAGGGTTATTTTATTTCTCAGC 80360

QY 121 TACTTAGTATTTTGTGGAGAGGGGTGACACCCAGACTAATGTCTTCATATGATC 180
 Db 80359 TACTTAGTATTTTGTGGAGAGGGGTGACACCCAGACTAATGTCTTCATATGATC 80300

QY 181 CTCATCGGAGGATTCAGATGATATGGACAGAAATCATCTAGGAGGACAAATGAACTTCA 240
 Db 80299 CTCATCGGAGGATTCAGATGATATGGACAGAAATCATCTAGGAGGACAAATGAACTTCA 80240


```
QY 241 ATGCAGATTCTAGCCACGCTAAAGATGAGAACACAGAAATCGCTGAAAACCTCTATTAG 299
|||||
Db 80239 ATGCAGATTCTAGCCACGCTAAAGATGAGAACACAGACATCGCTGAAAACCTCTATTAG 80181
|||||

RESULT 4
AC016638/c 182823 bp DNA linear PRI 21-JUL-2001
LOCUS Homo sapiens chromosome 5 clone RP11-365H8, complete sequence.
AC016638
AC016638.9 GI:14993683
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (18-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
REFERENCE
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Jul 21, 2001 this sequence version replaced gi:14861720.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.5% of Sequence;
Estimated Total Number of Errors is 0.7.
STS Content:
SHGC-3092 G16892
SHGC-18259 G19195
WI-4513 G02898.
Location/Qualifiers
1. 182823
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-365H8"
BASE COUNT 53032 a 39003 c 37540 g 53248 t
ORIGIN
Query Match 94.9%; Score 283.8; DB 9; Length 182823;
Best Local Similarity 99.0%; Pred. No. 1.8e-63;
Matches 296; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 1 GCAAGAGGAGCGCAGCGCGATGACAGAAAATTCACCTTCGCGAGATGCCCTCTAAATCCT 60
|||||
Db 121704 GCAAGAGGAGCGCAGCGCGATGACAG-AAAATACACTTTCGCGAGATGCCCTCTAAATCCT 121646
|||||
QY 61 GCCTCAATTTTAACTTCTCTATGATCGGCAATAGGGTTTATTTATTTCTTCAGC 120
|||||
Db 121645 GCTCAATTTTAACTTCTCTATGATCGGCAATAGGGTTTATTTATTTCTTCAGC 121586
|||||
QY 121 TACTTAGTATTTTGTGGAGAGAGGTTGACACCCAGACTAATGCTTCTTCATATATGATC 180
|||||
Db 121585 TACTTAGTATTTTGTGGAGAGAGGTTGACACCCAGACTAATGCTTCTTCATATATGATC 121526
|||||
QY 181 CTCATCGGAGGCATTTCAGATGATGATGAGACAGATCATCTAGGAGGACAAATGAACITCA 240
|||||

Db 121525 CTCATCGGAGGCATTTCAGATGATGATGAGACAGATCATCTAGGAGGACAAATGAACITCA 121466
|||||
QY 241 ATGCAGATTCTAGCCACGCTAAAGATGAGAACACAGAAATCGCTGAAAACCTCTATTAG 299
|||||
Db 121465 ATGCAGATTCTAGCCACGCTAAAGATGAGAACACAGACATCGCTGAAAACCTCTATTAG 121407
|||||

RESULT 5
AC079469 217807 bp DNA linear HTG 19-APR-2001
LOCUS Homo sapiens chromosome 5 clone RP11-490H18, WORKING DRAFT
DEFINITION SEQUENCE, 4 ordered pieces.
AC079469
AC079469.2 GI:13677063
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Apr 19, 2001 this sequence version replaced gi:9964834.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 594711
Center clone name: RPCI-11_490H18
-----
Summary Statistics
Consensus quality: 213920 bases at least Q40
Consensus quality: 216734 bases at least Q30
Consensus quality: 217279 bases at least Q20
Estimated insert size: 196000; pulse field gel estimation
Estimated insert size: 217507; sum-of-contigs estimation
Quality coverage: 8.1 in Q20 bases; pulse field gel estimation
Quality coverage: 7.37 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 28216: contig of 28216 bp in length
* 28217 28316: gap of unknown length
* 28317 41458: contig of 13142 bp in length
* 41459 41558: gap of unknown length
* 41559 155568: contig of 114010 bp in length
* 155569 155668: gap of unknown length
* 155669 217807: contig of 62139 bp in length.
Location/Qualifiers
1. 217807
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-490H18"
BASE COUNT 63364 a 45033 c 45462 g 63648 t 300 others
ORIGIN
```

```
Query Match          94.9%; Score 283.8; DB 2; Length 217807;
Best Local Similarity 99.0%; Pred. No. 1.8e-63;
Matches 296; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 GCAGAGGGAGCCACGCCCGATGACAGAAATTCACATTCCTCCGAGATGGCCCTCTAAATCCT 60
Db 10597 GCAGAGGGAGCCACGCCCGATGACAG-AAATACATTCCTCCGAGATGGCCCTCTAAATCCT 10655
QY 61 GGCTGAATTTTAACTTCTCTATGATCGCAATAGGTTTATTTATTTTCTCAGC 120
Db 10656 GGCTGAATTTTAACTTCTCTATGATCGCAATAGGTTTATTTATTTTCTCAGC 10715
QY 121 TACTACTATTTTGGGAGAGAGGGTGCACCCAGACAACTAATGTTCTTCATAATGATC 180
Db 10716 TACTACTATTTTGGGAGAGAGGGTGCACCCAGACAACTAATGTTCTTCATAATGATC 10775
QY 181 CTATCGGAGGATTCAGATGATAATGACAGAGATCATCTAGGAGCAAAATGAACCTTCA 240
Db 10776 CTATCGGAGGATTCAGATGATAATGACAGAGATCATCTAGGAGCAAAATGAACCTTCA 10835
QY 241 ATGCAGATTCAGCCAACTGTAAGATGAGAACACAGAAATCGCTGAAACCTCTATTAG 299
Db 10836 ATGCAGATTCAGCCAACTGTAAGATGAGAACACAGAAATCGCTGAAACCTCTATTAG 10894

RESULT 6
AX175160
LOCUS          AX175160          1794 bp      DNA          linear          PAT 03-JUL-2001
DEFINITION    Sequence 2 from Patent WO0144478.
ACCESSION     AX175160
VERSION       AX175160.1 GI:14598562
KEYWORDS
SOURCE        human.
ORGANISM      Homo sapiens
REFERENCE     1 (bases 1 to 1794)
AUTHORS      Canfield,W.M., Cummings,R.D. and Ju,T.
TITLE        Core lb3-galactosyl transferases and methods of use thereof
JOURNAL       Patent: WO 0144478-A 2 21-JUN-2001;
              The Board of Regents of The University of Oklahoma (US)
FEATURES
  source      1..1794
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /map="7p14-pl3"
              /chromosome="7"
              /note="unnamed protein product"
  CDS         63..1154
              /codon_start=1
              /protein_id="CAC43870.1"
              /db_xref="GI:14598563"
              /translation="MASKSWLNFLTFLCGSAIGFLCGLSOLFLLGKVDTPNVLHN
              DPARKSDNGONHLEGOMNFNADSSOHKDENTDIAENLYQKVRILCWMTGPNLEK
              KAKHVATWAQRCNKVLFWSSEENKDFPAVGLATKEGRQLYWKTITKAQYVHEHYLE
              DADWFLKADDDTYVILDNRLWLLSKYDPEPIYFGRFRFPYKVGKMSGAGYVLSKE
              ALKRFVDATKTDKTHSSIEDIALGRMEIMNVEAGDSRDITGKTFHPFVPEHHLI
              KGYLPRFWNYNYPVVEGPGCCSDLAVSFHYVDSVTMYELEYLVYHLRPGYLYR
              YOPTLPERILKEISQANKNEDTKVKLGNP"
  BASE COUNT  624 a 300 c 334 g 536 t
  ORIGIN
Query Match          79.1%; Score 236.4; DB 6; Length 1794;
Best Local Similarity 90.6%; Pred. No. 4.7e-51;
Matches 252; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 22 TGACAGAAATTCACATTCGAGATGGCCCTCTAAATCCTGGCTGAATTTTAACTTCC 81
Db 40 TTACAGAAATTCACATTCGAGAAATGGCCCTCTAAATCCTGGCTGAATTTTAACTTCC 99
QY 82 TCTATGATCGCAATAGGTTTATTTTATTTCTCAGCTACTAGTATTTTCTGGGAG 141
Db 100 TCTGTGATCAGCAATAGGATTTCTTTATGTTCTCAGCTATTAGTATTTCTGGGAG 159
QY 142 AAGAGGTGACACCCAGACAAATGTTCTTCAATATGATCCTCATGCGAGGCATTCAGATG 201
Db 142 AAGAGGTGACACCCAGCCTAAATGTTCTTCAATATGATCCTCATGCGAGGCATTCAGATG 219
QY 202 ATATGACAGAAATCATCTATGAGGAGGACAAATGAACCTCAATGATCTTAGCCAACTA 261
Db 220 ATATGACAGAAATCATCTATGAGGAGGACAAATGAACCTCAATGATCTTAGCCAACTA 279
QY 262 AAGATGAGAACACAGAAATCGCTGAAACCTCTATTAG 299
Db 280 AAGATGAGAACACAGAAATGCTGAAACCTCTATCAG 317

RESULT 7
AF155582
LOCUS          AF155582          1794 bp      mRNA          linear          PRI 03-JAN-2002
DEFINITION    Homo sapiens corel UDP-galactose:N-acetylgalactosamine-alpha-R beta
              1,3-galactosyltransferase (CIGALT1) mRNA, complete cds.
ACCESSION     AF155582
VERSION       AF155582.1 GI:8927161
KEYWORDS
SOURCE        Homo sapiens
ORGANISM      Homo sapiens
REFERENCE     1 (bases 1 to 1794)
AUTHORS      Ju,T., Brewer,K., D'Souza,A., Cummings,R.D. and Canfield,W.M.
TITLE        Cloning and Expression of Human Core 1 beta
              1,3-galactosyltransferase
JOURNAL       J. Biol. Chem. 277 (1), 178-186 (2002)
PUBMED        11677243
AUTHORS      Ju,T.Z., Cummings,R.D. and Canfield,W.M.
TITLE        Cloning and characterization of human corel
              UDP-galactose:N-acetylgalactosamine-alpha-R beta
              1,3-galactosyltransferase (Corel beta3-Gal-T) cDNA sequence
              Unpublished
JOURNAL       Direct Submission
AUTHORS      Submitted (01-JUN-1999) Department of Medicine, The University of
              Oklahoma, Health Science Center, 975 N.E. 10th Street, Biomedical
              Research Center, Room 411, Oklahoma City, OK 73104, USA
FEATURES
  source      1..1794
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /map="7p14-pl3"
              /chromosome="7"
              /note="corel beta3-Gal-T"
              /codon_start=1
              /product="corel
              UDP-galactose:N-acetylgalactosamine-alpha-R beta
              1,3-galactosyltransferase"
              /protein_id="AAF81981.1"
              /db_xref="GI:8927162"
              /translation="MASKSWLNFLTFLCGSAIGFLCGLSOLFLLGKVDTPNVLHN
              DPARKSDNGONHLEGOMNFNADSSOHKDENTDIAENLYQKVRILCWMTGPNLEK
              KAKHVATWAQRCNKVLFWSSEENKDFPAVGLATKEGRQLYWKTITKAQYVHEHYLE
              DADWFLKADDDTYVILDNRLWLLSKYDPEPIYFGRFRFPYKVGKMSGAGYVLSKE
              ALKRFVDATKTDKTHSSIEDIALGRMEIMNVEAGDSRDITGKTFHPFVPEHHLI
              KGYLPRFWNYNYPVVEGPGCCSDLAVSFHYVDSVTMYELEYLVYHLRPGYLYR
              YOPTLPERILKEISQANKNEDTKVKLGNP"
  misc_feature 75..158
              /gene="CIGALT1"
              /note="Region: non-cleavable transmembrane signal"
              /evidence=not_experimental
```

```
BASE COUNT      624 a   300 c   334 g   536 t
ORIGIN
ALKRFVDAFKTDKTHSSIEDLALGRCEIMNVVAGDSRDTIGKTFHPFVPEHHLI
KGYLPRTFYWYNNYPPVEVSLILLCCQYLD"
Query Match      79.1%; Score 236.4; DB 9; Length 1794;
Best Local Similarity 90.6%; Pred. No. 4.7e-51;
Matches 252; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
Qy 22 TGACAGAAATTCACCTTCCGAGATGGCCCTCTAAATCCTGGCTGAATTTTAAACCTCC 81
Db 40 TTACAGAAATACACTTTCGGGAAATGGCCCTAAATCCTGGCTGAATTTTAAACCTCC 99
Qy 82 TCTATGATCGGCAATAGGCTTATTTTCTCTCAGCTACTTACTAGTATTTTGTGGGAG 141
Db 100 TCTGTGATCAGCAATAGGATTTCTTTTATGTTCTCAGCTATTTAGTATTTTGTGGGAG 159
Qy 142 AAGAGGCTGACACCCAGCACTAAATGTTCTTCAATATGATCCTCATCGCAGGCATTCAGATG 201
Db 160 AAAAGGTTGACACCCAGCCCTAATGTTCTTCAATATGATCCTCATCGAAGCATTCAGATG 219
Qy 202 ATATGACAGATCATCTCTAGGAGGACAAATGAACCTTCAATGATGATCTTAGCCAAAGTA 261
Db 220 ATATGACAGATCATCTCTAGGAGGACAAATGAACCTTCAATGATGATCTTAGCCAAAGTA 279
Qy 262 AAGATGAGACAGCAATCGCTCAAAACCTCTATTAG 299
Db 280 AAGATGAGACAGCAATCGCTCAAAACCTCTATTAG 317

RESULT 8
HSA132443      930 bp   mRNA   linear   PRI 27-JUL-2001
LOCUS Homo sapiens mRNA for beta-1,3-galactosyltransferase b3Gal-T8.
ACCESSION AJ132443
VERSION AJ132443.1 GI:15028815
KEYWORDS b3Gal-T8; b3Gal-T8 gene; beta-1,3-galactosyltransferase.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Anado, M., Almeida, R., Schwientek, T. and Clausen, H.
IDENTIFICATION AND CHARACTERIZATION OF LARGE GALACTOSYLTRANSFERASE
JOURNAL gene families: galactosyltransferases for all functions
MEDLINE Biochim. Biophys. Acta 1473 (1), 35-53 (1999)
PUBMED 20047730
REFERENCE 10580128
AUTHORS Jensen, M.A. and Bennett, E.P.
TITLE Cloning of a new member of the beta-1,3-galactosyltransferase
JOURNAL family, b3Gal-T8
REFERENCE Unpublished
AUTHORS Jensen, M.A.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-1999) Jensen M.A., School of Dentistry,
AUTHORS University of Copenhagen, Nørre Alle 20, 2200 Copenhagen N, DENMARK
FEATURES
Location/Qualifiers
1..930
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="7p14-p13"
1..930
/gene="b3Gal-T8"
1..930
/gene="b3Gal-T8"
/codon_start=1
/product="beta-1,3-galactosyltransferase b3Gal-T8"
/protein_id="CAC45046.1"
/db_xref="GI:15028816"
/translation="MASKSWLFLTLFCGSAIGLLCSOLFSLICEKVDTPQNLHN
DPHARSDDNQNLHSGQMFNADSSQHKDNDIAENLQKVRILCWMTGQNLK
KAKHVRATQRCNKVLFMSSSEENKDFPVLKTKRGDQLYKRTIKAFQYVHEHYLE
DADWFLKADDDTYVILDNLRLSKYDPEEPIYFGRFRKPYVRQGYMSGGAGYVLSKE"

BASE COUNT      363 a   195 c   219 g   315 t
ORIGIN
ATGCGCTCTAAATCCTGGCTGAATTTTAACTTCTCTATGGATCGGCAATAGGTTT 104
Db 1 ATGCGCTCTAAATCCTGGCTGAATTTTAACTTCTCTATGGATCGGCAATAGGTTT 60
Qy 105 ATTTTATTTTCTCAGCTACTTAGTATTTTGTGGGAGAAAGGTTGACACCCAGACTAAT 164
Db 61 CTTTATGTTCTCAGCTACTTAGTATTTTGTGGGAGAAAGGTTGACACCCAGCTAAT 120
Qy 165 GTTCTTCATATGATCCTCATCGCAGGATTCAGATGATATGACAGAGATCATCTAGGA 224
Db 121 GTTCTTCATATGATCCTCATCGCAGGATTCAGATGATATGACAGAGATCATCTAGGA 180
Qy 225 GGACAAATGAACCTTCAATGATGATTCAGCAAGCATTCAGATGATATGACAGAGATCATCTAGGA 284
Db 181 GGACAAATGAACCTTCAATGATGATTCAGCAAGCATTCAGCAAGCATTCAGCAAGATTCAGCAAGATTCCT 240
Qy 285 GAAACCTCTATTAG 299
Db 241 GAAACCTCTATTAG 255

RESULT 9
AXI175170      1092 bp   DNA   linear   PAT 03-JUL-2001
LOCUS Sequence 12 from Patent WO0144478.
ACCESSION AXI175170
VERSION AXI175170.1 GI:14598568
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 1092)
JOURNAL Canfield, W.M., Cummings, R.D. and Ju, T.
MEDLINE Core lb3-galactosyl transferases and methods of use thereof
PUBMED Patent: WO 0144478-A 12 21-JUN-2001;
REFERENCE The Board of Regents of The University of Oklahoma (US)
FEATURES
Location/Qualifiers
1..1092
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT      363 a   195 c   219 g   315 t
ORIGIN
ATGCGCTCTAAATCCTGGCTGAATTTTAACTTCTCTATGGATCGGCAATAGGTTT 104
Db 1 ATGCGCTCTAAATCCTGGCTGAATTTTAACTTCTCTATGGATCGGCAATAGGTTT 60
Qy 105 ATTTTATTTTCTCAGCTACTTAGTATTTTGTGGGAGAAAGGTTGACACCCAGACTAAT 164
Db 61 CTTTATGTTCTCAGCTACTTAGTATTTTGTGGGAGAAAGGTTGACACCCAGCTAAT 120
Qy 165 GTTCTTCATATGATCCTCATCGCAGGATTCAGATGATATGACAGAGATCATCTAGGA 224
Db 121 GTTCTTCATATGATCCTCATCGCAGGATTCAGATGATATGACAGAGATCATCTAGGA 180
Qy 225 GGACAAATGAACCTTCAATGATGATTCAGCAAGCATTCAGATGATATGACAGAGATCATCTAGGA 284
Db 181 GGACAAATGAACCTTCAATGATGATTCAGCAAGCATTCAGCAAGCATTCAGCAAGATTCAGCAAGATTCCT 240
```

```

Qy 285 GAAACCTCTATTAG 299
|||||
Db 241 GAAACCTCTATCAG 255

RESULT 10
HSA243256 1092 bp mRNA linear PRI 04-JAN-2002
LOCUS Homo sapiens mRNA for beta-1,3-galactosyltransferase.
DEFINITION
ACCESSION AJ243256
VERSION AJ243256.1 GI:18073134
KEYWORDS b3Gal-T gene; beta-1,3-galactosyltransferase.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Jensen, M.P.A.
TITLE Cloning and expression of a novel beta-1,3-Galactosyltransferase
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1092)
AUTHORS Jensen, M.P.A.
TITLE Direct Submission
JOURNAL Submitted (23-JUN-1999) Jensen M.P.A., School of Dentistry,
University of Copenhagen, Noerre Alle 20, DK-2200 Copenhagen N,
DENMARK
FEATURES
source Location/Qualifiers
1..1092
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7p14-p13"
1..1092
/gene="b3Gal-T"
1..1092
/gene="b3Gal-T"
/function="making the T-antigen"
/codon_start=1
/product="beta-1,3-Galactosyltransferase"
/protein_id="CAC80435.1"
/db_xref="GI:18073135"
/translation="MASKSWLNFLLFCGSAIGLLCSQLFSILLGKVDTPQPNVLHN
DPHARHSDNGQNHLEQMFNADSSQHKDENTDIAENLYQKVRILCWVMTGPNLEK
KAKHVATQAQRCKNVLFMSSEENKDFPVLGKTKGRDQLYWTKTIFAFQVYVHEHYLE
DADFLKADDDTYVILDNLRLWLLSKYDPEEPIYFGRFPKPVKGYMGAGVYVLSKE
ALKRFVDAFTDKCTHSSIEDLALGRCEIMNVAGDSRDTIGKETFFHFPVEHHLI
KGYLPRTFWNTNYPVPEGPGCCSLAVSFHYVDSTTMYELEYLVHLRPYGYLYR
YQPLPERILKEISOANKNEDTKVLGNP"
BASE COUNT 363 a 195 c 219 g 315 t
ORIGIN
Query Match 77.3%; Score 231; DB 9; Length 1092;
Best Local Similarity 94.1%; Pred. No. 1.2e-49;
Matches 240; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 45 ATGCCCTCTAAATCCTCGCTGAATTTTAACTTCCTCTATGATCGGCAATAGGCTTT 104
|||||
Db 1 ATGCCCTCTAAATCCTCGCTGAATTTTAACTTCCTCTCTGATCAGCAATAGGATTT 60
|||||
Qy 105 ATTTTATTTCTCAGCTACTAGTATTGTTGGGAGAGAGGTGACACCCAGCTAAT 164
|||||
Db 61 CTTTATATGTTCTCAGCTATTAGTATTGTTGGGAGAGAGGTGACACCCAGCTAAT 120
|||||
Qy 165 GTTCTTCATATGATCTCATCGGAGCATTTCAGATGATATGACAGAAATCATCTAGGA 224
|||||
Db 121 GTTCTTCATATGATCTCATCGGAGCATTTCAGATGATATGACAGAAATCATCTAGGA 180
|||||
Qy 225 GGACAAATGAACCTTCAATTCAGATTTCTAGCCAACTTAAAGATGAGAACACAGAAATCGCT 284
|||||
Db 181 GGACAAATGAACCTTCAATTCAGATTTCTAGCCAACTTAAAGATGAGAACACAGATTTGCT 240
|||||
Qy 285 GAAACCTCTATTAG 299
|||||
Db 241 GAAACCTCTATCAG 255

RESULT 11
HSA278960 1092 bp mRNA linear PRI 05-DEC-2001
LOCUS Homo sapiens mRNA for beta 1,3-galactosyltransferase (core-1-beta
DEFINITION
ACCESSION AJ278960
VERSION AJ278960.1 GI:17384685
KEYWORDS beta 1,3-galactosyltransferase; core-1-beta
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Bennett, E.P.
TITLE Unpublished
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1092)
AUTHORS Bennett, E.P.
TITLE Direct Submission
JOURNAL Submitted (06-SEP-2000) Bennett E.P., Histo.lab., 24.5.33, School
of Dentistry, Noerre Alle' 20, 2200N, Copenhagen, DENMARK
FEATURES
source Location/Qualifiers
1..1092
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1092
/gene="core-1-beta 1,3-galactosyltransferase"
1..1092
/gene="core-1-beta 1,3-galactosyltransferase"
/function="Core-1 synthase"
/codon_start=1
/evidence=experimental
/product="beta 1,3-galactosyltransferase"
/protein_id="CAC82373.1"
/db_xref="GI:17384686"
/translation="MASKSWLNFLLFCGSAIGLLCSQLFSILLGKVDTPQPNVLHN
DPHARHSDNGQNHLEQMFNADSSQHKDENTDIAENLYQKVRILCWVMTGPNLEK
KAKHVATQAQRCKNVLFMSSEENKDFPVLGKTKGRDQLYWTKTIFAFQVYVHEHYLE
DADFLKADDDTYVILDNLRLWLLSKYDPEEPIYFGRFPKPVKGYMGAGVYVLSKE
ALKRFVDAFTDKCTHSSIEDLALGRCEIMNVAGDSRDTIGKETFFHFPVEHHLI
KGYLPRTFWNTNYPVPEGPGCCSLAVSFHYVDSTTMYELEYLVHLRPYGYLYR
YQPLPERILKEISOANKNEDTKVLGNP"
BASE COUNT 363 a 195 c 219 g 315 t
ORIGIN
Query Match 77.3%; Score 231; DB 9; Length 1092;
Best Local Similarity 94.1%; Pred. No. 1.2e-49;
Matches 240; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 45 ATGCCCTCTAAATCCTCGCTGAATTTTAACTTCCTCTATGATCGGCAATAGGCTTT 104
|||||
Db 1 ATGCCCTCTAAATCCTCGCTGAATTTTAACTTCCTCTCTGATCAGCAATAGGATTT 60
|||||
Qy 105 ATTTTATTTCTCAGCTACTAGTATTGTTGGGAGAGAGGTGACACCCAGCTAAT 164
|||||
Db 61 CTTTATATGTTCTCAGCTATTAGTATTGTTGGGAGAGAGGTGACACCCAGCTAAT 120
|||||
Qy 165 GTTCTTCATATGATCTCATCGGAGCATTTCAGATGATATGACAGAAATCATCTAGGA 224
|||||
Db 121 GTTCTTCATATGATCTCATCGGAGCATTTCAGATGATATGACAGAAATCATCTAGGA 180
|||||
Qy 225 GGACAAATGAACCTTCAATTCAGATTTCTAGCCAACTTAAAGATGAGAACACAGAAATCGCT 284
|||||
Db 181 GGACAAATGAACCTTCAATTCAGATTTCTAGCCAACTTAAAGATGAGAACACAGATTTGCT 240
|||||
Qy 285 GAAACCTCTATTAG 299
|||||
Db 241 GAAACCTCTATCAG 255

```

RESULT 12

AC005532
 LOCUS Homo sapiens PAC clone RP4-733B9 from 7p14-p13, complete sequence.
 DEFINITION AC005532
 ACCESSION AC005532
 VERSION AC005532.2 GI:21322201
 KEYWORDS HTGS.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 115954)
 Sulston, J.E. and Waterston, R.

Toward a complete human genome sequence
 Genome Res. 8 (11), 1097-1108 (1998)

99063792
 PUBMED
 9847074

2 (bases 1 to 115954)

Dauphin, S. and Gregory, S.

The sequence of Homo sapiens PAC clone RP4-733B9

Unpublished (2001)

3 (bases 1 to 115954)

Waterston, R.H.

Direct Submission

Submitted (25-AUG-1998) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

4 (bases 1 to 115954)

Waterston, R.

Direct Submission

Submitted (14-JAN-1999) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 5 (bases 1 to 115954)
 Waterston, R.

Direct Submission

Submitted (21-DEC-1999) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 6 (bases 1 to 115954)

Waterston, R.

Direct Submission

Submitted (04-JUN-2002) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Jun 4, 2002 this sequence version replaced gi:4156192.

COMMENT

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: sapiens@wustl.edu

----- Summary Statistics

----- Center project name: H_DU0733B09

NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by sequence
 from more than one subclone; and the assembly was confirmed by
 restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and
 sequencing collaboration between the NHGRI Chromosome 7 Mapping
 Project (Eric D. Green, Director), John D. McPherson in the
 Department of Genetics (Washington University), and the Washington
 University Genome Sequencing Center. For additional information
 about the map position of this sequence, see
<http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send

<mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-4, prepared by
 Pieter de Jong and coworkers at <http://www.chori.org> using the
 method described by Ioannou et al., Nature Genetics 6:84-9 (1994).
 The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc.

(<http://www.genomesystems.com>) or Research Genetics, Inc.

(<http://www.resgen.com>); or from Pieter de Jong.

VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

The actual start of this clone is at base position 1 of RP4-733B9,
 actual end is at 115954 of RP4-733B9.

Base positions 26118 to 26479 and 25639 to 25745 of RP4-733B9 are
 covered by per product from RP4-733B9 clone DNA.

FEATURES
 Source

1. 115954

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/map="7p14-p13"

/clone="RP4-733B9"

/clone_lib="RPCI-4"

1. 221

/rpt_family="L1"

repeat_region

222. 520

/rpt_family="Alu"

repeat_region

521. 2344

/rpt_family="L1"

repeat_region

2364. 2774

/rpt_family="L1"

repeat_region

2775. 3195

/rpt_family="MaLR"

repeat_region

3196. 3297

/rpt_family="L1"

repeat_region

3298. 3342

/rpt_family="MaLR"

repeat_region

3344. 3630

/rpt_family="Alu"

repeat_region

3646. 3993

/rpt_family="MaLR"

repeat_region

3998. 4184

/rpt_family="L1"

repeat_region

4185. 4316

/rpt_family="Alu"

repeat_region

4317. 4361

/rpt_family="L1"

repeat_region

4357. 5291

/rpt_family="L1"

repeat_region

5392. 5686

/rpt_family="MaLR"

repeat_region

5671. 5890

/rpt_family="MER21-group"

repeat_region

7540. 7699

/rpt_family="MIR"

repeat_region

8200. 8227

/rpt_family="AT-rich"

repeat_region

8458. 8518

/rpt_family="L2"

repeat_region

8467. 8525

/rpt_family="MIR"

repeat_region

8531. 9081

/rpt_family="MaLR"

repeat_region

9397. 9829

/rpt_family="MaLR"

repeat_region

10026. 10197

/rpt_family="L1"

repeat_region

10465. 10645

/rpt_family="Alu"

repeat_region

10791. 10822

repeat_region

```
repeat_region /rpt_family="AT-rich"
repeat_region /rpt_family="MIR"
repeat_region /rpt_family="MIR"
misc_feature 12111..13064
repeat_region /note="CpG island (%GC=73.3, o/e=0.90, #CpGs=112)"
repeat_region /rpt_family="MaLR"
repeat_region /rpt_family="L2"
repeat_region 15851..15864
repeat_region /rpt_family="MaLR"
repeat_region 15865..15888
repeat_region /rpt_family="(TTTC)n"
repeat_region 15889..16044
repeat_region /rpt_family="MaLR"
repeat_region 16045..16130
repeat_region /rpt_family="L2"
repeat_region 16131..16278
repeat_region /rpt_family="MERL_type"
repeat_region 16279..16495
repeat_region /rpt_family="L2"
repeat_region 16502..16560
repeat_region /rpt_family="Mariner"
repeat_region 16568..16849
repeat_region /rpt_family="L2"
repeat_region 16876..17173
repeat_region /rpt_family="Alu"
repeat_region 17231..17412
repeat_region /rpt_family="MERL_type"
repeat_region 17440..18164
repeat_region /rpt_family="L2"
repeat_region 18304..18324
repeat_region /rpt_family="AT-rich"
repeat_region 19772..20224
repeat_region /rpt_family="LTR65"
repeat_region 20438..20473
repeat_region /rpt_family="AT-rich"
repeat_region 20689..21860
repeat_region /rpt_family="L1"
repeat_region 21872..25747
repeat_region /rpt_family="L1"
repeat_region 25748..25778
repeat_region /rpt_family="(TTG)n"
repeat_region 25779..25794
repeat_region /rpt_family="L1"
repeat_region 25795..31297
repeat_region /rpt_family="L1"
repeat_region 31249..32108
repeat_region /rpt_family="L1"
repeat_region 32101..33423
repeat_region /rpt_family="L1"
repeat_region 33424..33700
repeat_region /rpt_family="L1"
repeat_region 33701..34935

Query Match 69.18; Score 206.6; DB 9; Length 115954;
Best Local Similarity 90.28; Pred. No. 2e-43;
Matches 221; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Qy 22 TGACAGAAAATTCACCTTTCCGAGATGCCCTCTAAATCCTGGCTGAATTTTTTAACCTTC 81
Db 64018 TTACAGAAATACACTTCGGGAATGCCCTCTAAATCCTGGCTGAATTTTTTAACCTTC 64077

Qy 82 TCATGATCGGCAATAGGCTTTATTTTCTCAGCTACTAGTATTTGTTGGAG 141
Db 64078 TCCTGATGATCAGCAATAGGATTTCTTTATGTTCTCAGCTATTTAGTATTTTGTGGAG 64137

Qy 142 AAGAGGTGACACCCAGCAATGCTCTTCATATGATCTCCTGCGGCAATTCAGATG 201
Db 64138 AAAGGTTGACACCCAGCCTAATGTTCTTCATATGATCTCCTGCAAGCATTCAGATG 64197
```

```
Qy 202 ATAATGACAGAAATCATCTAGGAGGACAAATGAACCTTCAATGAGATTTCTAGCCACGTA 261
Db 64198 ATAATGACAGAAATCATCTAGGAGGACAAATGAACCTTCAATGAGATTTCTAGCCACATA 64257

Qy 262 AAGAT 266
Db 64258 AAGGT 64262

RESULT 13
AF157962 1438 bp mRNA linear ROD 03-JAN-2002
LOCUS Mus musculus corel UDP-galactose:N-acetylgalactosamine-alpha-R beta
DEFINITION 1,3-galactosyltransferase (Cigalt1) mRNA, complete cds.
ACCESSION AF157962
VERSION AF157962.1 GI:8927163
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1438)
AUTHORS Ju, T. Z., Brewer, K., D'Souza, A., Cummings, R. D. and Canfield, W. M.
TITLE Cloning and Expression of Human Core 1 beta
JOURNAL 1,3-Galactosyltransferase
PUBMED 11677243
REFERENCE 2 (bases 1 to 1438)
AUTHORS Ju, T. Z., Cummings, R. D. and Canfield, W. M.
TITLE Cloning and characterization of corel
JOURNAL UDP-galactose:N-acetylgalactosamine-alpha-R beta
REFERENCE 3 (bases 1 to 1438)
AUTHORS Ju, T. Z., Cummings, R. D. and Canfield, W. M.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-1999) Department of Medicine, The University of
Oklahoma, Health Science Center, 975 N.E. 10th Street, Biomedical
Research Center, Room 411, Oklahoma City, OK 73104, USA

FEATURES
source 1. .1438
/organism="Mus musculus"
/db_xref="taxon:10090"
1. .1438
/gene="Cigalt1"
149..1240
/gene="Cigalt1"
/EC_number="2.4.1.122"
/function="responsible for the synthesis of corel
structure of O-glycans"
/note="corel beta3-Gal-T"
/codon_start=1
/product="corel
UDP-galactose:N-acetylgalactosamine-alpha-R beta
1,3-galactosyltransferase"
/protein_id="AAF81982.1"
/db_xref="GI:8927164"
/translation="MASKSWLNFVLCSSAIGFFLCSSQLLSILLREERAIQPNMLHN
DPHARSDHNGSHLKGQMNFNADSSQHKDENIDVAENLYQVKVILCWNTSPQNLK
KAKHVATQORCNKYLFMSSSEKQDPTVGLKTKREGQLYKTKIKAFQYVHDHYLE
DADWFKADDDTIVIDNLRLLSKYNPEQPIYFGRFRFPYVKQYMSGAGYVLSKE
ALRRFVNATKTEKTHSSSTIEDLALGRMEIINVEAGDSRDTIGKETFFHPVPEHHLI
KGYLPKTFWNNYVPPIEGGCCSDIAVSFHYVDGTTMYELEYLVYHLRPGYLYR
YQPALDENILKEINQVNRKDKIKLGNP"
misc_feature 161..241
/gene="Cigalt1"
/note="Region: non-cleavable transmembrane signal"
/evidence=not_experimental

BASE COUNT 465 a 286 c 298 g 389 t
ORIGIN

Query Match 64.9%; Score 194; DB 10; Length 1438;
Best Local Similarity 84.5%; Pred. No. 4.7e-40;
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 22:29:21 ; Search time 117.627 Seconds
(without alignments)
5724.434 Million cell updates/sec

Title: US-10-029-359A-1

Perfect score: 299

Sequence: 1 gcaagaggagccacggcg.....tcgtgaaacacctattag 299

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|-------|---------------|--------|----------|---------------------|
| 1 | 298 | 99.7 | 505 | ABK88407 | Human DNA encoding |
| 2 | 236.4 | 79.1 | 1794 | AAF90507 | Human core 1 beta3 |
| 3 | 236.4 | 79.1 | 1800 | AAZ99551 | Nucleotide sequenc |
| 4 | 224.2 | 75.0 | 1424 | AAZ73202 | DNA encoding novel |
| 5 | 194 | 64.9 | 1469 | AAF90509 | Mouse core 1 beta3 |
| 6 | 189.2 | 63.3 | 1440 | AAF90508 | Rat core 1 beta3-g |
| 7 | 91 | 30.4 | 1560 | ABK88409 | Human DNA encoding |
| 8 | 66.2 | 22.1 | 795 | AAL24100 | Human breast cancer |
| 9 | 46.2 | 15.5 | 489 | ABK88408 | Human DNA encoding |

| | | | | | | |
|----|------|------|--------|----|----------|--------------------|
| 10 | 37.6 | 12.6 | 2646 | 16 | AAT02405 | Rice mature pullul |
| 11 | 37.6 | 12.6 | 2982 | 16 | AAT02400 | Pullulanase expres |
| 12 | 37.6 | 12.6 | 2988 | 16 | AAT02399 | Rice mature pullul |
| 13 | 36 | 12.0 | 4282 | 21 | AAC48551 | Arabidopsis thalia |
| 14 | 35.6 | 11.9 | 10859 | 22 | AAL04720 | Human reproductiv |
| 15 | 35.6 | 11.9 | 10859 | 23 | ABL97627 | Human testicular a |
| 16 | 35.4 | 11.8 | 1083 | 23 | AAS76745 | DNA encoding novel |
| 17 | 35.2 | 11.8 | 513 | 22 | AAL15254 | Human breast,cance |
| 18 | 34.2 | 11.4 | 4265 | 21 | AAH29796 | S cerevisiae apopt |
| 19 | 34 | 11.4 | 342 | 22 | AAH44815 | Human secreted exp |
| 20 | 34 | 11.4 | 4764 | 22 | ABA08912 | Human cancer assoc |
| 21 | 34 | 11.4 | 4764 | 22 | AAH90024 | Human bone marrow |
| 22 | 34 | 11.4 | 4767 | 22 | AAH89911 | Human cancer assoc |
| 23 | 34 | 11.4 | 5807 | 21 | AAA09157 | Human cancer assoc |
| 24 | 34 | 11.4 | 6380 | 21 | AAA09159 | Drosophila melanog |
| 25 | 34 | 11.4 | 11314 | 23 | ABL06868 | Human immune/haema |
| 26 | 34 | 11.4 | 17666 | 22 | AAK76302 | Human ovarian PCR- |
| 27 | 33.8 | 11.3 | 618 | 22 | AAS23915 | DNA encoding novel |
| 28 | 33.8 | 11.3 | 618 | 22 | AAH82472 | Human ovarian tumo |
| 29 | 33.8 | 11.3 | 1782 | 23 | AAH78516 | DNA encoding novel |
| 30 | 33.8 | 11.3 | 2001 | 24 | ABK34618 | Human cDNA for nov |
| 31 | 33.8 | 11.3 | 2263 | 23 | AAS92449 | DNA encoding novel |
| 32 | 33.8 | 11.3 | 18692 | 23 | ABK42795 | Genomic sequence # |
| 33 | 33.6 | 11.2 | 3647 | 23 | ABL18680 | Drosophila melanog |
| 34 | 33.6 | 11.2 | 4153 | 23 | ABL12654 | Drosophila melanog |
| 35 | 33.4 | 11.2 | 6935 | 24 | ABL92262 | Chemically treated |
| 36 | 33.4 | 11.2 | 47066 | 23 | ABL11514 | Drosophila melanog |
| 37 | 33.4 | 11.2 | 349980 | 22 | AAH41225 | Pyrococcus abyssi |
| 38 | 33.2 | 11.1 | 2123 | 20 | AAH20304 | Borrelia burgdorfe |
| 39 | 33.2 | 11.1 | 26241 | 22 | ABA16222 | Human nervous syst |
| 40 | 33.2 | 11.1 | 147708 | 24 | ABQ88154 | Human osteoblast d |
| 41 | 33 | 11.0 | 837 | 22 | AAH53653 | S. epidermidis ope |
| 42 | 33 | 11.0 | 3236 | 22 | AAH54128 | S. epidermidis gen |
| 43 | 33 | 11.0 | 9145 | 24 | ABL32888 | Human immune syste |
| 44 | 32.8 | 11.0 | 1707 | 22 | AAH45215 | Human zinc finger |
| 45 | 32.8 | 11.0 | 1919 | 22 | AAK03010 | Human brain expres |

ALIGNMENTS

| | | |
|----------|----------|--|
| RESULT 1 | ABK88407 | standard; DNA; 505 BP. |
| ID | ABK88407 | |
| XX | AC | ABK88407; |
| XX | DT | 07-OCT-2002 (first entry) |
| XX | DE | Human DNA encoding partial protein cluster II protein #1. |
| XX | KW | Human; ds; gene; protein cluster II; obesity; diabetes mellitus; |
| XX | KW | central nervous system disorder; metabolic disease. |
| XX | OS | Homo sapiens. |
| XX | FH | Key |
| XX | FT | CDS |
| XX | FT | Location/Qualifiers |
| XX | FT | 21..497 |
| XX | FT | /*tag= |
| XX | FT | /product= "protein cluster II protein #1" |
| XX | FT | /note= "No stop codon shown" |
| XX | FT | /transl_except= (pos:297..299,aa:aa) |
| XX | FT | /note= "Xaa= unknown" |
| XX | PN | WO200251864-A1. |
| XX | XX | |
| XX | PD | 04-JUL-2002. |
| XX | XX | |
| XX | PF | 14-DEC-2001; 2001WO-SE02786. |
| XX | XX | |
| XX | PR | 22-DEC-2000; 2000SE-0004828. |
| XX | XX | |
| XX | PA | (PHAA) PHARMACIA AB. |

XX PI Attersand A;
XX DR WPI: 2002-575368/61.
XX DR P-PSDB; ABG30851.
XX PT New protein cluster II nucleic acids and polypeptides, useful in
XX PT diagnosing metabolic diseases such as obesity and diabetes, and in
XX PT identifying agents for treating such diseases
XX PS Claim 1; Page 20-21; 28pp; English.
XX CC The invention relates to a new isolated nucleic acid comprising:
XX CC (a) a fully defined sequence, encoding a protein cluster II protein,
XX CC appearing as ABK8407 and ABK8408; (b) a sequence capable of hybridising
XX CC under stringent hybridisation conditions to a nucleotide sequence
XX CC complementary to the polypeptide coding region of a nucleic acid in (a);
XX CC or (c) a sequence which is degenerate as a result of the genetic code to
XX CC a nucleotide sequence in (a) or (b). Also included are a isolated
XX CC an polypeptide encoded by the nucleic acid, a vector harbouring the
XX CC nucleic acid, a replicable expression vector, which carries and is
XX CC capable of mediating the expression of the nucleic acid a cultured host
XX CC cell harbouring the expression vector, producing a polypeptide by
XX CC culturing the cell, where the polypeptide is produced, and recovering the
XX CC polypeptide and identifying an agent capable of modulating the nucleic
XX CC acid by providing a cell comprising the nucleic acid molecule, contacting
XX CC the cell with a candidate agent, and monitoring the cell for an effect
XX CC that is not present in the absence of the candidate agent. The protein
XX CC cluster II nucleic acid and polypeptide is useful in the diagnosis of
XX CC metabolic diseases such as obesity and diabetes, and central nervous
XX CC system disorders and in the identification of agents for treating these
XX CC diseases. The nucleic acids may be used as hybridisation probes, for
XX CC chromosome and gene mapping, in polymerase chain reaction (PCR)
XX CC technologies, in the production of sense and antisense nucleic acids, and
XX CC in screening for new therapeutic molecules. The present sequence
XX CC encodes a protein cluster II protein of the invention.
XX SQ Sequence 505 BP; 166 A; 95 C; 108 G; 135 T; 1 other;
Query Match 99.7%; Score 298; DB 24; Length 505;
Best Local Similarity 99.7%; Pred. No. 9.4e-81;
Matches 298; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GCAAGAGGAGCGACGGCGATGACAGAAAATTCACCTTCCGAGATGGCTCTAAATCCT 60
Db 1 GCAAGAGGAGCGACGGCGATGACAGAAAATTCACCTTCCGAGATGGCTCTAAATCCT 60
QY 61 GGCTCAATTTTAACTTCTCTATGATCGGCAATAGGGTTTATTTATTTCTCAGC 120
Db 61 GGCTCAATTTTAACTTCTCTATGATCGGCAATAGGGTTTATTTATTTCTCAGC 120
QY 121 TACTTAGTATTGTTGGGAGAGGGGTGACACCCAGACTAATCTTCTCATATGATC 180
Db 121 TACTTAGTATTGTTGGGAGAGGGGTGACACCCAGACTAATCTTCTCATATGATC 180
QY 181 CTCATCGAGGATTCAGATGATATGACAGAAATCATCTAGGAGGACAAATGAACCTCA 240
Db 181 CTCATCGAGGATTCAGATGATATGACAGAAATCATCTAGGAGGACAAATGAACCTCA 240
QY 241 ATGCAGATTCAGCAACGTAAGATGAGACACAGAAATCGCTGAAACCTCTATTAG 299
Db 241 ATGCAGATTCAGCAACGTAAGATGAGACACAGAAATCGCTGAAACCTCTATTAG 299
RESULT 2
ID AAF90507
XX AAF90507 standard; cDNA; 1794 BP.
XX AC AAF90507;
XX XX
XX 22-AUG-2001 (first entry)
XX DE Human core 1 beta3-galactosyl transferase cDNA.

XX KW Core 1 beta3-galactosyl transferase; human; O-glycosylation;
XX KW galactosylation; glycosulfopeptide; Tn syndrome; IgA nephropathy;
XX KW diagnosis; therapy; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT 63..1154
XX FT /*tag= a
XX FT /note= "a polynucleotide comprising this
XX FT coding sequence is also claimed in Claim
XX FT 1(A)"
XX PN WO200144478-A2.
XX PD 21-JUN-2001.
XX PF 14-DEC-2000; 2000WO-US33945.
XX PR 15-DEC-1999; 99US-0461321.
XX PR 15-DEC-1999; 99US-0464035.
XX PA (UYOK-) UNIV OKLAHOMA.
XX PI Canfield WM, Cummings RD, Ju T;
XX WI: 2001-398157/42.
XX P-PSDB; AAB82456.
XX DR Novel purified core 1 beta3 galactosyl transferases of human, rat,
XX DR mouse, Drosophila melanogaster, Caenorhabditis elegans, useful for
XX DR galactosylating N-acetyl galactosamine linked to O-linking amino acid
XX DR on protein
XX PS Claim 1(A); Fig 3; 74pp; English.
XX CC The present sequence is that of human cDNA for core 1
XX CC beta3-galactosyltransferase (b3GTase, see AAB82456). The coding
XX CC region of the sequence is also claimed. The cDNA was identified
XX CC initially as an expressed sequence tag (EST) following BlastN
XX CC searching with a rat EST sequence, and was completed by 5'-RACE
XX CC using human placenta cDNA as template. The encoded protein is a
XX CC type 2 transmembrane protein. The invention provides human, rat,
XX CC mouse, Drosophila melanogaster and Caenorhabditis elegans b3GTases
XX CC and nucleic acids encoding them. The polypeptides exhibit a wide
XX CC range of homologies. The polynucleotides can be used to transform
XX CC or transfect host cells for producing substantially pure forms of
XX CC the enzyme, or for use in an expression system for core 1 O-linked
XX CC glycosylation of proteins or peptides produced within the
XX CC expression system. The expressed enzymes galactosylate, via a
XX CC beta3 linkage, an N-acetyl galactosamine linked to a serine,
XX CC threonine or other O-linking amino acid on peptides or proteins.
XX CC Core 1 b3GTases are useful in the synthesis of glycosulfopeptides
XX CC which can function as inhibitors of P-selectin/PSGL-1 interactions.
XX CC Other potential uses include diagnostic tests for the rare
XX CC Tn-syndrome or IgA nephropathy, and the therapy of these disorders.
XX SQ Sequence 1794 BP; 624 A; 300 C; 334 G; 536 T; 0 other;

Query Match 79.1%; Score 236.4; DB 22; Length 1794;
Best Local Similarity 90.6%; Pred. No. 8.3e-62;
Matches 252; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY 22 TCACAGAAATTCACCTTCCGAGATGGCTCTAAATCTCGCTGAATTTTAACTTCC 81
Db 40 TTACAGAAATACACTTTCGGGAATGGCTCTAAATCTCGCTGAATTTTAACTTCC 99
QY 82 TCTATGGATCGCAATAGGTTTATTTATTTTCTCAGCTACTAGTATTTGTTGGAG 141
Db 100 TCTGTGGATCAGCAATAGGATTTCTTTATGTTCTCAGCTATTAGTATTTGTTGGAG 159
QY 142 AAGAGGGTGACACCCAGACTAATGTTCTTCATATGATCTCATCGGAGGCAATTCAGATG 201

160 AAAAGGTTGACACCCAGCCTAATGTTCTCATAATGATCTCTCAAGGCAATTCAGATG 219
202 ATAATGACAGAAATCATCTAGGAGGACAAATGAATCAATGAGATTCAGTCCACGTA 261
220 ATAATGACAGAAATCATCTAGGAGGACAAATGAATCAATGAGATTCAGTCCACGTA 279
262 AAGATGAGACACAGAAATCGCTGAAACCTCTATTAG 299
280 AAGATGAGACACAGAAATCGCTGAAACCTCTATTAG 317

RESULT 3

AZ299551
ID AZ299551 standard; DNA; 1800 BP.

XX AC AZ299551;

XX DT 03-JUL-2000 (first entry)

XX DE Nucleotide sequence of human core-1 2 beta1,3-GalT.

XX KW Glycosulfopeptide; P-selectin glycoprotein ligand-1;
XX KW inflammatory disease; chronic inflammation; acute inflammation;
XX KW diffuse inflammation; traumatic inflammation; immunosuppression;
XX KW toxic diffuse inflammation; specific inflammation; core-1 2 beta1,3-GalT;
XX KW reactive inflammation; parenchymatous inflammation;
XX KW obliterative inflammation; interstitial inflammation;
XX KW croupous inflammation; focal inflammation; rheumatoid arthritis;
XX KW acute leukocyte-mediated lung injury; acute inflammation; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 63..1154

XX FT /*tag= a

XX FT /product= "core-1 2 beta1,3-GalT"

XX PN WO9965712-A2.

XX PD 23-DEC-1999.

XX PF 15-JUN-1999; 99WO-US13455.

XX PR 16-JUN-1998; 98US-0089472.

XX PA (CUMM//) CUMMINGS R D.

XX PA (MCEV//) MCEVER R P.

XX PI Cummings RD, McEver RP;

XX DR WPI; 2000-237146/20.

XX DR P-PSDB; AAY84165.

XX PT Novel synthetic glycosulfopeptides which are P-selectin glycoprotein
XX PT ligand-1 (PSGL-1) mimics, used as anti-inflammatory agents in the
XX PT treatment of acute and chronic inflammation

XX PS Example 1; Page 73-75; 80pp; English.

XX CC The present sequence encodes a human core-1 2 beta1,3-GalT polypeptide.
XX CC The polypeptide is used to produce the synthetic glycosulfopeptides of
XX CC the invention. The specification describes a new class of synthetic
XX CC glycosulfopeptides which mimic the extreme amino terminus of P-selectin
XX CC glycoprotein ligand-1. The glycosulfopeptides comprise one or more
XX CC sulphated tyrosine residues and a glycan comprising a sialyl Lewis-x
XX CC group or a sialyl Lewis-a group. The synthetic glycosulfopeptides are
XX CC used to treat inflammatory diseases. They can be used to treat both
XX CC chronic and acute inflammatory diseases. e.g. diffuse inflammation, traumatic
XX CC inflammation, immunosuppression, toxic diffuse inflammation, specific
XX CC inflammation, reactive inflammation, parenchymatous inflammation, obliterative
XX CC inflammation, interstitial inflammation, croupous
XX CC inflammation and focal inflammation. Diseases which may be treated

CC include rheumatoid arthritis, post-ischemic (reperfusion)
CC leukocyte-mediated tissue damage, acute leukocyte-mediated lung
CC injury (e.g. Adult respiratory Distress syndrome), and other tissue or
CC organ specific forms of acute inflammation (e.g. glomerulonephritis).
CC The GSPs can also be used with enzyme linked immunosorbant assay (ELISA)
CC techniques to distinguish between monoclonal antibodies which react with
CC core-2 sialyl Lewis-x groups versus those which react with core-1 sialyl
CC Lewis-x groups. The glycosulfopeptides are also excellent acceptors for
XX specific glycosyltransferases.

XX SQ Sequence 1800 BP; 625 A; 302 C; 336 G; 537 T; 0 other;

Query Match 79.1%; Score 236.4; DB 21; Length 1800;

Best Local Similarity 90.6%; Pred. No. 8.3e-62;

Matches 252; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 22 TGACAGAAAATTCACCTTCCGAGATGCGCTCTAAATCCTGGCTGAATTTTAACTTCC 81

Db 40 TTACAGAAATACACTTTCGGGAATGGCTCTAAATCCTGGCTGAATTTTAACTTCC 99

QY 82 TCTATGGATCGCAATAGGGTTTATTTATTTCTCAGCTACTTAGTATTTTGGGAG 141

Db 100 TCTGTGGATCAGCAATAGGATTTCTTTATCTCTCAGCTATTTAGTATTTTGGGAG 159

QY 142 AAGAGGGTGACACCCAGACTAATGTTCTTCTATATGATCTCTCGAGGCAATTCAGATG 201

Db 160 AAGAGGGTGACACCCAGACTAATGTTCTTCTATATGATCTCTCGAGGCAATTCAGATG 219

QY 202 ATAATGGACAGAAATCATCTAGGAGGACAAATGAATTCATGAGATCTAGGCAACGTA 261

Db 220 ATAATGGACAGAAATCATCTAGGAGGACAAATGAATTCATGAGATCTAGGCAACGTA 279

QY 262 AAGATGAGACACAGAAATCGCTGAAAACCTCTATTAG 299

Db 280 AAGATGAGACACAGAAATCGCTGAAAACCTCTATTAG 317

RESULT 4

AAS73202

ID AAS73202 standard; cDNA; 1424 BP.

XX AC AAS73202;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #9006.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSPO INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR P-PSDB; ABG09015.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -

Claim 1; SEQ ID No 9006; 103pp; English.

PS The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome CC and gene mapping, and in recombinant production of (II). The CC polynucleotides are also used in diagnostics as expressed sequence tags CC for identifying expressed genes. (I) is useful in gene therapy techniques CC to restore normal activity of (II) or to treat disease states involving CC (II). (II) is useful for generating antibodies against it, detecting or CC quantitating a polypeptide in tissue, as molecular weight markers and as CC a food supplement. (II) and its binding partners are useful in medical CC imaging of sites expressing (II). (I) and (II) are useful for treating CC disorders involving aberrant protein expression or biological activity. CC The polypeptide and polynucleotide sequences have applications in CC diagnostics, forensics, gene mapping, identification of mutations CC responsible for genetic disorders or other traits to assess biodiversity CC and to produce other types of data and products dependent on DNA and CC amino acid sequences. AAS64197-AAS94564 represent novel human CC diagnostic coding sequences of the invention. CC Note: The sequence data for this patent did not appear in the printed CC specification, but was obtained in electronic format directly from WIPO CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1424 BP; 456 A; 285 C; 315 G; 368 T; 0 other;

Query Match 75.0%; Score 224.2; DB 23; Length 1424;
Best Local Similarity 92.9%; Pred. No. 3.9e-58;
Matches 235; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
QY 37 TTTCGAGATGGCTCTAAATCCTGGCTGAATTTTAACTTCTCTATGGATCGCAA 96
DB 9 TTTCGGAATGCGCTCTAAATCCTGGCTGAATTTTAACTTCTCTATGGATCGCAA 68
QY 97 TAGGCTTTATTTTCTAGCTACTAGTATTTTGGGAGAGAGGTGACACC 156
DB 69 TAGGATTTTCTTATGTTCTAGCTATTTAGTATTTTGGGAGAGAGGTGACACC 128
QY 157 AGACTAATGTTCTTCAATATGATCCTCATGCGAGGCTTCAGATCATATGACAGAATC 216
DB 129 AGCTAATGTTCTTCAATATGATCCTCATGCGAGGCTTCAGATCATATGACAGAATC 188
QY 217 ATCTAGGAGGCAAAATGAACCTTCAATGAGATTTAGCCAAACGTAAGATGAGAACACAG 276
DB 189 ATCTAGGAGGCAAAATGAACCTTCAATGAGATTTAGCCAAACGTAAGATGAGAACACAG 248
QY 277 AAATCGCTCAAAA 289
DB 249 ACATTGCTGAAAA 261

RESULT 5
AAAF90509
ID AAF90509 standard; cDNA; 1469 BP.

XX AAF90509;

XX 22-AUG-2001 (first entry)

XX .Mouse core 1 beta3-galactosyl transferase cDNA.

XX Core 1 beta3-galactosyl transferase; mouse; O-glycosylation;
KW galactosylation; glycosulfopeptide; Tn syndrome; IgA nephropathy;
KW diagnosis; therapy; ss;

OS Mus musculus.

XX Key Location/Qualifiers
FH 180..1271
CDS /*tag= a

FT /note= "a polynucleotide comprising this
FT coding sequence is also claimed in Claim
FT 1(A)."

XX WO200144478-A2.
XX 21-JUN-2001.
XX 14-DEC-2000; 2000WO-US33945.
XX 15-DEC-1999; 99US-0461321.
XX 15-DEC-1999; 99US-0464035.
XX (UYOK-) UNIV OKLAHOMA.
XX Canfield WM, Cummings RD, Ju T;
XX WPI; 2001-398157/42.
XX P-PSDB; AAB82458.
XX Novel purified core 1 beta3 galactosyl transferases of human, rat,
XX mouse, Drosophila melanogaster, Caenorhabditis elegans, useful for
XX galactosylating N-acetyl galactosamine linked to O-linking amino acid
XX on protein
XX Claim 1(A); Page 64; 74pp; English.
XX The present sequence is that of mouse cDNA for core 1
XX beta3-galactosyltransferase (b3GTase, see AAB82458). The mouse
XX core 1 b3GTase has 89% identity and 94% similarity to the human
XX enzyme (see AAB82456). The invention provides human, rat, mouse,
XX Drosophila melanogaster and Caenorhabditis elegans b3GTases and
XX nucleic acids encoding them. The polypeptides exhibit a wide
XX range of homologies. The polynucleotides can be used to transform
XX or transfect host cells for producing substantially pure forms of
XX the enzyme, or for use in an expression system for core 1 O-linked
XX glycosylation of proteins or peptides produced within the
XX expression system. The expressed enzymes can be used to
XX galactosylate, via a beta3 linkage, an N-acetylgalactosamine linked
XX to a serine, threonine or other O-linking amino acid on peptides or
XX proteins. Core 1 b3GTases are also useful in the synthesis of
XX glycosulfopeptides which can function as inhibitors of
XX P-selectin:PSGL-1 interactions. Other potential uses include
XX diagnostic tests for the rare Tn-syndrome or IgA nephropathy, and
XX therapy of these disorders.
XX Sequence 1469 BP; 457 A; 296 C; 313 G; 393 T; 0 other;
Query Match 64.9%; Score 194; DB 22; Length 1469;
Best Local Similarity 84.5%; Pred. No. 6.1e-49;
Matches 218; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
QY 42 GAGATGGCTCTAAATCCTGGCTGAATTTTAACTTCTCTATGGATCGCAATAGG 101
DB 177 GAATGGCTCTAAATCCTGGCTGAATTTTAACTTCTCTATGGATCGCAATAGG 236
QY 102 TTTATTTTATTTTCTCAGCTACTTAGTATTTTGGGAGAGAGGTGACACCCAGACT 161
DB 237 TTTTATTTATCTTCTCAACTCTTGAGTATTTTGGGAGAGAGGTGACACCCAGCT 296
QY 162 AATGTTCTTCATATGATCCTCATGCGAGGCTTCAGATCATATGACAGAATCATCTA 221
DB 297 AACATGCTTCAATGACCTCATGCAAGGATTCAGATGACATGGAACAGTCACTC 356
QY 222 GGAGCAAAATGAACCTTCAATGCGAGTCTAGCCCAACCTAAAGATGAGAACACAGAAATC 281
DB 357 AAGGACAGATGAACCTTCAATGCGAGTTCAGCCCAACATTAAGATGAGAACATAGACGTT 416
QY 282 GCTGAAAACCTCTATTAG 299
DB 417 GCTGAGAACCTCTATCAG 434
RESULT 6
AAAF90508
ID AAF90508 standard; cDNA; 1440 BP.

[illegible]

[illegible]

| | | | |
|-----------|-----|--------------|-----|
| Qy | 284 | TGAAACCTCTA | 295 |
| | | | |
| Db | 405 | TCTGGACCTCTA | 416 |
| RESULT 11 | | | |

AA02400
ID AAT02400 standard; DNA; 2982 BP.
XX
XX
AC AAT02400;
XX
XX
DT 30-JAN-1996 (first entry)
XX
XX
PR

| | |
|----|--|
| KW | Pullulanase; rice; brewing; low-calorie beer; high dextrose syrup; |
| KW | yeast; Saccharomyces cerevisiae; ds. |
| XX | |
| OS | Oryza sativa var. IR 36. |
| XX | |
| XX | Key |
| FH | Location/Qualifiers |
| FT | misc_difference 2901 |
| FT | /*tag= a |
| FT | /note= "base n at position 2901 is not identified |
| FT | in the specification" |
| XX | |
| PN | W09509922-Al. |
| XX | |

PD 13-APR-1995.
vv

PF 04-OCT-1994; 94WO-US11242.

PR 05-OCT-1993; 93US-0132648.

XX
PA (MILL-) MILLER BREWING CO.

XX
PI Bower PA:

XX
DR WPT: 1995-155263/20

XX
XX
DE

yeast, useful in the brewing and beverage industries

PS Claim 3; Page 23-24; 43pp; English.

CC 5' 147 bp and 3' 701 bp end fragments of the rice pullulanase gene

were combined with a 2.3 kb portion of pullulanase cDNA clone 6-1 (see AAT02399) to obtain an expression construct (AAT02400) encoding a

pullulanase lacking the first 2 amino acids of the native enzyme. The construct was used to express pullulanase in *E. coli*.

SEY2102 (ATCC 74281) for use in low-calorie beer or high dextrose

| | |
|--|---|
| | C |
| | E |
| | F |
| | G |
| | H |
| | I |
| | J |
| | K |
| | L |
| | M |
| | N |
| | O |
| | P |
| | Q |
| | R |
| | S |
| | T |
| | U |
| | V |
| | W |
| | X |
| | Y |
| | Z |
| | A |
| | B |
| | C |
| | D |
| | E |
| | F |
| | G |
| | H |
| | I |
| | J |
| | K |
| | L |
| | M |
| | N |
| | O |
| | P |
| | Q |
| | R |
| | S |
| | T |
| | U |
| | V |
| | W |
| | X |
| | Y |
| | Z |
| | A |
| | B |
| | C |
| | D |
| | E |
| | F |
| | G |
| | H |
| | I |
| | J |
| | K |
| | L |
| | M |
| | N |
| | O |
| | P |
| | Q |
| | R |
| | S |
| | T |
| | U |
| | V |
| | W |
| | X |
| | Y |
| | Z |
| | A |
| | B |
| | C |
| | D |
| | E |
| | F |
| | G |
| | H |
| | I |
| | J |
| | K |
| | L |
| | M |
| | N |
| | O |
| | P |
| | Q |
| | R |
| | S |
| | T |
| | U |
| | V |
| | W |
| | X |
| | Y |
| | Z |
| | A |
| | B |
| | C |
| | D |
| | E |
| | F |
| | G |
| | H |
| | I |
| | J |
| | K |
| | L |
| | M |
| | N |
| | O |
| | P |
| | Q |
| | R |
| | S |
| | T |
| | U |
| | V |
| | W |
| | X |
| | Y |
| | Z |
| | A |
| | B |
| | C |
| | D |
| | E |
| | F |
| | G |
| | H |
| | I |
| | J |
| | K |
| | L |
| | M |
| | N |
| | O |
| | P |
| | Q |
| | R |
| | S |
| | T |
| | U |
| | V |
| | W |
| | X |
| | Y |
| | Z |
| | A |
| | B |
| | C |
| | D |
| | E |
| | F |
| | G |
| | H |
| | I |
| | J |
| | K |
| | L |
| | M |
| | N |
| | O |
| | P |
| | Q |
| | R |
| | S |
| | T |
| | U |
| | V |
| | W |
| | X |
| | Y |
| | Z |
| | A |
| | B |
| | C |
| | D |
| | E |
| | F |
| | G |
| | H |
| | I |
| | J |
| | K |
| | L |
| | M |
| | N |
| | O |
| | P |
| | Q |
| | R |
| | S |
| | T |
| | U |
| | V |
| | W |
| | X |
| | Y |
| | Z |
| | A |
| | B |
| | C |
| | D |
| | E |
| | F |
| | G |
| | H |
| | I |
| | J |
| | K |
| | L |
| | M |
| | N |
| | O |
| | P |
| | Q |
| | R |
| | S |
| | T |
| | U |
| | V |
| | W |
| | X |
| | Y |
| | Z |
| | A |
| | B |
| | C |
| | D |
| | E |
| | F |
| | G |
| | H |
| | I |
| | J |
| | K |
| | L |
| | M |
| | N |
| | O |
| | P |
| | Q |
| | R |
| | S |
| | T |
| | U |
| | V |
| | W |
| | X |
| | Y |
| | Z |
| | A |
| | B |
| | C |
| | D |
| | E |
| | F |
| | G |
| | H |
| | I |
| | J |
| | K |
| | L |
| | M |
| | N |
| | O |
| | P |
| | Q |
| | R |
| | S |
| | T |
| | U |
| | V |
| | W |
| | X |
| | Y |
| | Z |
| | A |
| | B |
| | C |
| | D |
| | E |
| | F |

22. Sequence 2502 BT, 075 A, 004 C, 030 G, 002 I, 1 Other;

| | | | | |
|-----------------------|--------|-----------------|--------|--------------|
| Query MACCN | 12.6%; | Score 37.6; | DB 16; | Length 2982; |
| Best Local Similarity | 55.3%; | Pred. No. 0.31; | | |

```
Matches 73; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
```

OY 164 TGTTCCTCATTAAGATCCTCATCGGAGGCATTCAGATGATAATGACAGAAATCATCTAGG 223
 DB 279 TGTGCTCTTATGATGCTCATGGGAGCGTCAAGATGTTACTGGATTGCAACTACCTGG 338
 OY 224 AGGACAATGAACTTCATGAGATTCAGGATTCAGCCATGTAAGATGACAGAAATCGC 283
 DB 339 TGTATTGGATGACATGTTGCTTATGCTTACTGGACCACTTGGTGCAGTTCAGTGATAAAGA 398
 OY 284 TGAACACCTCTA 295
 DB 399 TGTGGACCTCTA 410
 RESULT 12
 AAT02399
 ID AAT02399 standard; CDNA; 2988 BP.
 XX
 AC AAT02399;
 XX
 DT 30-JAN-1996 (first entry)
 XX
 DE Rice mature pullulanase gene.
 XX
 KW Pullulanase; rice; brewing; low-calorie beer; high dextrose syrup;
 KW yeast; Saccharomyces cerevisiae; ds.
 XX
 OS Oryza sativa var. IR 36.
 XX
 FH Key Location/Qualifiers
 FT misc_difference 2907
 FT /*tag= a
 FT /note= "base n at position 2907 is not identified
 in the specification"
 PN - WO9509922-A1.
 XX
 PD 13-APR-1995.
 XX
 PF 04-OCT-1994; 94WO-US11242.
 XX -
 PR 05-OCT-1993; 93US-01326648.
 XX (MILL-) MILLER BREWING CO.
 PA
 XX Bower PA;
 XX
 DR WPI; 1995-155263/20.
 XX
 FT Constructs contg. cloned rice pullulanase gene - for expression in
 FT yeast, useful in the brewing and beverage industries
 XX
 PS Disclosure; Page 21-22; 43pp; English.
 XX
 CC A rice flowering stage cDNA library was screened using a rice
 CC pullulanase genomic clone and a probe based on the pullulanase N-
 CC terminal sequence. cDNA clone 6-1 was obtd. This encoded a
 CC pullulanase lacking 13 N-terminal amino acids of the mature
 CC protein. The DNA sequence for these missing amino acids was
 CC determined from a PCR fragment and from a genomic clone. The
 CC resulting sequence for the mature pullulanase gene is given in
 CC AAT02399.
 XX
 SQ Sequence 2988 BP; 876 A; 554 C; 695 G; 862 T; 1 other;
 Query Match 12.68; Score 37.6; DB 16; Length 2988;
 Best Local Similarity 55.38; Pred. No. 0.31; Indels 0; Gaps 0;
 Matches 73; Conservative 0; Mismatches 59;
 OY 164 TGTTCCTCATTAAGATCCTCATCGGAGGCATTCAGATGATAATGACAGAAATCATCTAGG 223
 DB 285 TGTGCTCTTATGATGCTCATGGGAGCGTCAAGATGTTACTGGATTGCAACTACCTGG 344
 OY 224 AGGACAATGAACTTCATGAGATTCAGGATTCAGCCATGTAAGATGACAGAAATCGC 283

DB 345 TGTATTGGATGACATGTTGCTTATGCTTATGCTGACCACTTGGTGCAGTTCAGTGATAAAGA 404
 OY 284 TGAACACCTCTA 295
 DB 405 TGTGGACCTCTA 416
 RESULT 13
 AAC48551
 ID AAC48551 standard; DNA; 4282 BP.
 XX
 AC AAC48551;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 57889.
 XX
 KW Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 23-MAR-1999; 99US-0126284.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 18-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.

| | | | | | |
|----|--------------|---------------|----|--------------|---------------|
| PR | 16-JUN-1999; | 99US-0139452. | PR | 12-AUG-1999; | 99US-0148341. |
| PR | 16-JUN-1999; | 99US-0139453. | PR | 13-AUG-1999; | 99US-0148565. |
| PR | 17-JUN-1999; | 99US-0139492. | PR | 13-AUG-1999; | 99US-0148684. |
| PR | 18-JUN-1999; | 99US-0139454. | PR | 16-AUG-1999; | 99US-0149358. |
| PR | 18-JUN-1999; | 99US-0139455. | PR | 17-AUG-1999; | 99US-0149175. |
| PR | 18-JUN-1999; | 99US-0139456. | PR | 18-AUG-1999; | 99US-0149426. |
| PR | 18-JUN-1999; | 99US-0139457. | PR | 20-AUG-1999; | 99US-0149722. |
| PR | 18-JUN-1999; | 99US-0139458. | PR | 20-AUG-1999; | 99US-0149723. |
| PR | 18-JUN-1999; | 99US-0139459. | PR | 20-AUG-1999; | 99US-0149929. |
| PR | 18-JUN-1999; | 99US-0139460. | PR | 23-AUG-1999; | 99US-0149902. |
| PR | 18-JUN-1999; | 99US-0139461. | PR | 23-AUG-1999; | 99US-0149930. |
| PR | 18-JUN-1999; | 99US-0139462. | PR | 25-AUG-1999; | 99US-0150566. |
| PR | 18-JUN-1999; | 99US-0139463. | PR | 26-AUG-1999; | 99US-0150884. |
| PR | 18-JUN-1999; | 99US-0139750. | PR | 27-AUG-1999; | 99US-0151065. |
| PR | 18-JUN-1999; | 99US-0139763. | PR | 27-AUG-1999; | 99US-0151066. |
| PR | 21-JUN-1999; | 99US-0139817. | PR | 27-AUG-1999; | 99US-0151080. |
| PR | 22-JUN-1999; | 99US-0139899. | PR | 30-AUG-1999; | 99US-0151303. |
| PR | 23-JUN-1999; | 99US-0140353. | PR | 31-AUG-1999; | 99US-0151438. |
| PR | 23-JUN-1999; | 99US-0140354. | PR | 01-SEP-1999; | 99US-0151930. |
| PR | 24-JUN-1999; | 99US-0140695. | PR | 07-SEP-1999; | 99US-0152363. |
| PR | 28-JUN-1999; | 99US-0140823. | PR | 10-SEP-1999; | 99US-0153070. |
| PR | 29-JUN-1999; | 99US-0140991. | PR | 13-SEP-1999; | 99US-0153758. |
| PR | 30-JUN-1999; | 99US-0141287. | PR | 15-SEP-1999; | 99US-0154018. |
| PR | 01-JUL-1999; | 99US-0141842. | PR | 16-SEP-1999; | 99US-0154039. |
| PR | 01-JUL-1999; | 99US-0142154. | PR | 20-SEP-1999; | 99US-0154779. |
| PR | 02-JUL-1999; | 99US-0142055. | PR | 22-SEP-1999; | 99US-0155139. |
| PR | 06-JUL-1999; | 99US-0142390. | PR | 23-SEP-1999; | 99US-0155486. |
| PR | 08-JUL-1999; | 99US-0142803. | PR | 24-SEP-1999; | 99US-0155659. |
| PR | 09-JUL-1999; | 99US-0142920. | PR | 28-SEP-1999; | 99US-0156458. |
| PR | 12-JUL-1999; | 99US-0142977. | PR | 29-SEP-1999; | 99US-0156596. |
| PR | 13-JUL-1999; | 99US-0143542. | PR | 04-OCT-1999; | 99US-0157117. |
| PR | 14-JUL-1999; | 99US-0143624. | PR | 05-OCT-1999; | 99US-0157753. |
| PR | 15-JUL-1999; | 99US-0144005. | PR | 06-OCT-1999; | 99US-0157865. |
| PR | 16-JUL-1999; | 99US-0144085. | PR | 07-OCT-1999; | 99US-0158029. |
| PR | 16-JUL-1999; | 99US-0144086. | PR | 08-OCT-1999; | 99US-0158232. |
| PR | 19-JUL-1999; | 99US-0144325. | PR | 12-OCT-1999; | 99US-0158369. |
| PR | 19-JUL-1999; | 99US-0144331. | PR | 13-OCT-1999; | 99US-0159293. |
| PR | 19-JUL-1999; | 99US-0144332. | PR | 13-OCT-1999; | 99US-0159294. |
| PR | 19-JUL-1999; | 99US-0144333. | PR | 13-OCT-1999; | 99US-0159295. |
| PR | 19-JUL-1999; | 99US-0144334. | PR | 14-OCT-1999; | 99US-0159329. |
| PR | 19-JUL-1999; | 99US-0144335. | PR | 14-OCT-1999; | 99US-0159330. |
| PR | 20-JUL-1999; | 99US-0144352. | PR | 14-OCT-1999; | 99US-0159331. |
| PR | 20-JUL-1999; | 99US-0144632. | PR | 14-OCT-1999; | 99US-0159637. |
| PR | 20-JUL-1999; | 99US-0144884. | PR | 14-OCT-1999; | 99US-0159638. |
| PR | 21-JUL-1999; | 99US-0144814. | PR | 18-OCT-1999; | 99US-0159584. |
| PR | 21-JUL-1999; | 99US-0145086. | PR | 21-OCT-1999; | 99US-0160741. |
| PR | 21-JUL-1999; | 99US-0145088. | PR | 21-OCT-1999; | 99US-0160767. |
| PR | 22-JUL-1999; | 99US-0145085. | PR | 21-OCT-1999; | 99US-0160768. |
| PR | 22-JUL-1999; | 99US-0145087. | PR | 21-OCT-1999; | 99US-0160770. |
| PR | 22-JUL-1999; | 99US-0145089. | PR | 21-OCT-1999; | 99US-0160814. |
| PR | 22-JUL-1999; | 99US-0145192. | PR | 21-OCT-1999; | 99US-0160815. |
| PR | 23-JUL-1999; | 99US-0145145. | PR | 22-OCT-1999; | 99US-0160980. |
| PR | 23-JUL-1999; | 99US-0145218. | PR | 22-OCT-1999; | 99US-0160981. |
| PR | 23-JUL-1999; | 99US-0145224. | PR | 22-OCT-1999; | 99US-0160989. |
| PR | 26-JUL-1999; | 99US-0145276. | PR | 25-OCT-1999; | 99US-0161404. |
| PR | 27-JUL-1999; | 99US-0145913. | PR | 25-OCT-1999; | 99US-0161405. |
| PR | 27-JUL-1999; | 99US-0145918. | PR | 25-OCT-1999; | 99US-0161406. |
| PR | 27-JUL-1999; | 99US-0145919. | PR | 26-OCT-1999; | 99US-0161359. |
| PR | 28-JUL-1999; | 99US-0145951. | PR | 26-OCT-1999; | 99US-0161360. |
| PR | 02-AUG-1999; | 99US-0146386. | PR | 26-OCT-1999; | 99US-0161361. |
| PR | 02-AUG-1999; | 99US-0146388. | PR | 28-OCT-1999; | 99US-0161920. |
| PR | 02-AUG-1999; | 99US-0146389. | PR | 28-OCT-1999; | 99US-0161992. |
| PR | 03-AUG-1999; | 99US-0147038. | PR | 28-OCT-1999; | 99US-0161993. |
| PR | 04-AUG-1999; | 99US-0147204. | PR | 29-OCT-1999; | 99US-0162142. |
| PR | 04-AUG-1999; | 99US-0147302. | | | |
| PR | 05-AUG-1999; | 99US-0147192. | | | |
| PR | 05-AUG-1999; | 99US-0147260. | | | |
| PR | 06-AUG-1999; | 99US-0147303. | | | |
| PR | 06-AUG-1999; | 99US-0147416. | | | |
| PR | 09-AUG-1999; | 99US-0147493. | | | |
| PR | 09-AUG-1999; | 99US-0147935. | | | |
| PR | 10-AUG-1999; | 99US-0148171. | | | |
| PR | 11-AUG-1999; | 99US-0148319. | | | |

Query Match 12.0%; Score 36; DB 21; Length 4282;

Best Local Similarity 49.0%; Pred. No. 1.1;

Matches 96; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

Qy 85 ATGGATCGGCAATAGGGTTATTTATTTCTCAGCTACTTAGTATTTTGTGGGAGAAG 144

Db 809 ATGCATTGGACCAACCAATAGAGGGTTAGCTGAGTTCTTATGATCGTCTTGGAGATG 868

| | | | |
|-----------|---|--|------|
| Qy | 145 | AGGGTGACACCCAGACTAATGTTCTTCAATAATGATCTCATGGAGGCGATTCAGATGATA | 204 |
| | | | |
| Db | 869 | AGGCTAATTCATCGGCTCTCGAGATTTCTTAATGGTGATACTAAGTCGCAAGAACATGAAT | 928 |
| | | | |
| Qy | 205 | ATGGACAGAAATCATCTAGGAGGACAGAAATGAACCTTCAATGCGAGATTTAGCCACACGTAAG | 264 |
| | | | |
| Db | 929 | CGTGCACACTCAATTTTACAGCAACTTGGTTCTTGACAAACAAATCTCAAGGACACAGATG | 988 |
| | | | |
| Qy | 265 | ATGAGAACACAGAAAT | 280 |
| | | | |
| Db | 989 | ACGAGCTGACAGAAAT | 1004 |
| | | | |
| RESULT | 14 | | |
| AAAL04720 | | | |
| ID | AAAL04720 | standard; DNA; 10859 BP. | |
| XX | AAAL04720; | | |
| XX | XX | | |
| DT | 21-NOV-2001 | (first entry) | |
| DE | Human reproductive system related antigen DNA SEQ ID NO: 7408. | | |
| XX | Human; reproductive system related antigen; reproductive system disorder; | | |
| KW | cancer; gene therapy; ds. | | |
| XX | XX | | |
| OS | Homo sapiens. | | |
| XX | XX | | |
| PN | WO200155320-A2. | | |
| XX | XX | | |
| PD | 02-AUG-2001. | | |
| XX | XX | | |
| PF | 17-JAN-2001; 2001WO-US01339. | | |
| XX | XX | | |
| PR | 31-JAN-2000; 2000US-0179065. | | |
| PR | 04-FEB-2000; 2000US-0180628. | | |
| PR | 24-FEB-2000; 2000US-0184664. | | |
| PR | 02-MAR-2000; 2000US-0186350. | | |
| PR | 16-MAR-2000; 2000US-0189874. | | |
| PR | 17-MAR-2000; 2000US-0190076. | | |
| PR | 18-APR-2000; 2000US-0198123. | | |
| PR | 19-MAY-2000; 2000US-0205515. | | |
| PR | 07-JUN-2000; 2000US-0209467. | | |
| PR | 28-JUN-2000; 2000US-0214886. | | |
| PR | 30-JUN-2000; 2000US-0215135. | | |
| PR | 07-JUL-2000; 2000US-0216647. | | |
| PR | 07-JUL-2000; 2000US-0216880. | | |
| PR | 11-JUL-2000; 2000US-0217487. | | |
| PR | 11-JUL-2000; 2000US-0217496. | | |
| PR | 14-JUL-2000; 2000US-0218290. | | |
| PR | 26-JUL-2000; 2000US-0220963. | | |
| PR | 26-JUL-2000; 2000US-0220964. | | |
| PR | 14-AUG-2000; 2000US-0224518. | | |
| PR | 14-AUG-2000; 2000US-0224519. | | |
| PR | 14-AUG-2000; 2000US-0225213. | | |
| PR | 14-AUG-2000; 2000US-0225214. | | |
| PR | 14-AUG-2000; 2000US-0225266. | | |
| PR | 14-AUG-2000; 2000US-0225267. | | |
| PR | 14-AUG-2000; 2000US-0225268. | | |
| PR | 14-AUG-2000; 2000US-0225270. | | |
| PR | 14-AUG-2000; 2000US-0225447. | | |
| PR | 14-AUG-2000; 2000US-0225757. | | |
| PR | 14-AUG-2000; 2000US-0225758. | | |
| PR | 14-AUG-2000; 2000US-0225759. | | |
| PR | 18-AUG-2000; 2000US-0226279. | | |
| PR | 22-AUG-2000; 2000US-0226581. | | |
| PR | 22-AUG-2000; 2000US-0226868. | | |
| PR | 22-AUG-2000; 2000US-0227182. | | |
| PR | 23-AUG-2000; 2000US-0227009. | | |
| PR | 30-AUG-2000; 2000US-0228924. | | |
| PR | 01-SEP-2000; 2000US-0229287. | | |
| PR | 01-SEP-2000; 2000US-0229343. | | |
| PR | 01-SEP-2000; 2000US-0229344. | | |

| Result No. | Score | Query | | Length | DB | ID | Description |
|------------|-------|-------|--------|--------|---------------------|-------------------|-------------|
| | | Match | | | | | |
| 1 | 37.6 | 12.6 | 2646 | 1 | US-08-365-189-7 | Sequence 7, Appli | |
| 2 | 37.6 | 12.6 | 2982 | 1 | US-08-365-189-2 | Sequence 2, Appli | |
| 3 | 37.6 | 12.6 | 2988 | 1 | US-08-365-189-1 | Sequence 1, Appli | |
| 4 | 32.8 | 11.0 | 2042 | 2 | US-08-911-434A-2 | Sequence 2, Appli | |
| 5 | 32 | 10.7 | 800 | 4 | US-09-222-938A-68 | Sequence 68, Appl | |
| 6 | 31.8 | 10.6 | 1674 | 2 | US-08-713-410-31 | Sequence 31, Appl | |
| c 7 | 31.2 | 10.4 | 6216 | 4 | US-09-415-522-5 | Sequence 5, Appli | |
| c 8 | 31 | 10.4 | 1059 | 4 | US-08-837-199A-23 | Sequence 23, Appl | |
| c 9 | 31 | 10.4 | 1059 | 4 | US-08-837-199A-47 | Sequence 47, Appl | |
| 10 | 30.8 | 10.3 | 1506 | 1 | US-08-220-958-1 | Sequence 1, Appli | |
| c 11 | 30 | 10.0 | 3541 | 4 | US-09-180-439-5 | Sequence 5, Appli | |
| c 12 | 30 | 10.0 | 3979 | 4 | US-09-180-439-1 | Sequence 1, Appli | |
| c 13 | 30 | 10.0 | 3979 | 4 | US-09-180-439-2 | Sequence 2, Appli | |
| c 14 | 30 | 10.0 | 4123 | 4 | US-09-180-439-7 | Sequence 7, Appli | |
| c 15 | 30 | 10.0 | 6471 | 4 | US-09-353-585-1 | Sequence 1, Appli | |
| c 16 | 29.8 | 10.0 | 12311 | 4 | US-08-750-717-1 | Sequence 1, Appli | |
| 17 | 29.6 | 9.9 | 1371 | 4 | US-09-134-001C-935 | Sequence 935, App | |
| c 18 | 29.6 | 9.9 | 1893 | 4 | US-09-134-001C-1778 | Sequence 1778, Ap | |
| c 19 | 29.6 | 9.9 | 6816 | 4 | US-09-404-650-1 | Sequence 3, Appli | |
| c 20 | 29.6 | 9.9 | 6855 | 4 | US-09-404-650-3 | Sequence 3, Appli | |
| 21 | 29.6 | 9.9 | 37948 | 4 | US-09-251-645-11 | Sequence 11, Appl | |
| 22 | 29.4 | 9.8 | 1524 | 4 | US-09-134-001C-1141 | Sequence 1141, Ap | |
| c 23 | 29.4 | 9.8 | 4865 | 3 | US-08-894-017-24 | Sequence 24, Appl | |
| 24 | 29.4 | 9.8 | 8920 | 2 | US-08-446-855A-1 | Sequence 1, Appli | |
| 25 | 29.4 | 9.8 | 8920 | 4 | US-09-150-741-1 | Sequence 1, Appli | |
| c 26 | 29.4 | 9.8 | 111282 | 4 | US-09-754-250-3 | Sequence 3, Appli | |
| c 27 | 29.2 | 9.8 | 389 | 4 | US-09-370-838-105 | Sequence 105, App | |

| | | |
|-----|---|-----|
| 224 | AGGACAATGAACCTTCAATGCGAGATTCTACCCAAACGTAAGAATGAGAACACAGAAATGCG | 283 |
| Oy | | |
| 345 | TGTATTGGATGACATGTTTGCCTTATACCTGGACCACCTTGGTGCAGTTTTCAGTGTATAAGA | 404 |
| Db | | |
| 284 | TGAAAACCTCTA | 295 |
| Oy | | |
| 405 | TGTGGACCTCTA | 416 |
| Db | | |

RESULT 2

US-08-365-189-2
Sequence 2, Application US/08365189
Patent No. 5514576
GENERAL INFORMATION:
APPLICANT: Bower, Patricia A.
TITLE OF INVENTION: Cloned Pullulanase
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: U.S.A.
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patencin Release #1.0, Vers
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/365,189
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/132,648
FILING DATE: October 5, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Ryser, David G.
REGISTRATION NUMBER: 36,407
REFERENCE/DOCKET NUMBER: 66-005-9367-7
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5717
TELEFAX: (414) 271-3552
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2982 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Oligonucleotide
US-08-365-189-2

| | Query Match | 12.6% | Score 37.6; | DB 1; | Length 2982; |
|----|-----------------------|---|------------------|-------|----------------------------------|
| | Best Local Similarity | 55.3%; | Pred. No. 0.013; | | |
| | Matches | 73; | Conservative | 0; | Mismatches 59; Indels 0; Gaps 0; |
| QY | 164 | TGTTCTTCATATGATCCCTCATCGGAGGCAATTCAGATGATAATGGACAGCAATCATCTAGG | 223 | | |
| | | | | | |
| Db | 279 | TGTCGGCTCTTATGATGCTCATCGGGAGGCGTCAAGATGTTACTGGATTGGCAACTACCTGG | 338 | | |
| | | | | | |
| QY | 224 | AGGACAAATGAACCTTCAATGTCGACATCTTAGCCAACTGAAAGATGAGAAACACAGAAATTCGC | 283 | | |
| | | | | | |
| Db | 339 | TGTATTGGATGACATGTTGCTTATACTGGACCACCTGGTGCAGTTTTCAGTGATAAAGA | 398 | | |
| | | | | | |
| QY | 284 | TGAAACCTCTA | 295 | | |
| | | | | | |
| Db | 399 | TGTGGACCTCTA | 410 | | |
| | | | | | |

RESULT 3

US-08-365-189-1
; Sequence 1, Application US/08365189

```

1  ; Patent No. 5514576
2  ;
3  ; GENERAL INFORMATION:
4  ;
5  ; APPLICANT: Bower, Patricia A.
6  ;
7  ; TITLE OF INVENTION: Cloned Pullulanase
8  ;
9  ; NUMBER OF SEQUENCES: 14
10 ;
11 ; CORRESPONDENCE ADDRESS:
12 ;
13 ; ADDRESSEE: Quarles & Brady
14 ;
15 ; STREET: 411 East Wisconsin Avenue
16 ;
17 ; CITY: Milwaukee
18 ;
19 ; STATE: Wisconsin
20 ;
21 ; COUNTRY: U.S.A.
22 ;
23 ; ZIP: 53202-4497
24 ;
25 ; COMPUTER READABLE FORM:
26 ;
27 ; MEDIUM TYPE: Floppy disk
28 ;
29 ; COMPUTER: IBM PC compatible
30 ;
31 ; OPERATING SYSTEM: PC-DOS/MS-DOS
32 ;
33 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
34 ;
35 ; CURRENT APPLICATION DATA:
36 ;
37 ; APPLICATION NUMBER: US/08/365.189
38 ;
39 ; FILING DATE:
40 ;
41 ; CLASSIFICATION: 435
42 ;
43 ; PRIOR APPLICATION DATA:
44 ;
45 ; APPLICATION NUMBER: 08/132.648
46 ;
47 ; FILING DATE: October 5, 1993
48 ;
49 ; ATTORNEY/AGENT INFORMATION:
50 ;
51 ; NAME: Ryser, David G.
52 ;
53 ; REGISTRATION NUMBER: 36,407
54 ;
55 ; REFERENCE/DOCKET NUMBER: 66-005-9367-4
56 ;
57 ; TELECOMMUNICATION INFORMATION:
58 ;
59 ; TELEPHONE: (414) 277-5717
60 ;
61 ; TELEFAX: (414) 271-3552
62 ;
63 ; INFORMATION FOR SEQ ID NO: 1:
64 ;
65 ; SEQUENCE CHARACTERISTICS:
66 ;
67 ; LENGTH: 2988 base pairs
68 ;
69 ; TYPE: nucleic acid
70 ;
71 ; STRANDEDNESS: double
72 ;
73 ; TOPOLOGY: linear
74 ;
75 ; MOLECULE TYPE: cDNA
76 ;
77 ; PS-08-365.189-1

```

| | Query Match | 12.6% | Score 37.6; | DB 1; | Length 2988; | |
|-----|-----------------------|------------------------|-------------------|---------------------|----------------------------------|----------|
| | Best Local Similarity | 55.3%; | Pred. No. 0.013; | | | |
| | Matches | 73; | Conservative | 0; | Mismatches 59; Indels 0; Gaps 0; | |
| 164 | QY | TGTTCTTCAATATGATGACCTC | CGAGGCATTCAGATGAT | AATGGACAGAAATCATCT | PAGG 223 | |
| | | | | | | |
| 285 | Db | TTCGCGTCTTTATGCTGATTC | GGAGGGGCTAAGATGTT | TACTGGATTGCAAGTACCT | CG 344 | |
| | | | | | | |
| 224 | QY | AGGACAAATGAACTTCAAT | TGCAGATTC | TACGCCAACCT | TAAGATGGAGAACACAGAAAT | TCGC 283 |
| | | | | | | |
| 345 | Db | TGATTTGGATGACATGTTT | GCTTATATCGACCACT | TGGTGCAGTTTTCAGT | GATTAAGA 404 | |
| | | | | | | |

RESULT 4

US-08-911-434A-2
Sequence 2, Application US/08911434A
Patent No. 5959176
GENERAL INFORMATION:
APPLICANT: TORIKAI, Satomi
APPLICANT: OEDA, Kenji
TITLE OF INVENTION: PLANT PROMOTER AND UTILIZATION THEREOF
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP
STREET: P.O. BOX 747
CITY: FALLS CHURCH
STATE: VIRGINIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 22040

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911.434A
; FILING DATE: 12-AUG-1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Raymond C.
; REGISTRATION NUMBER: 21,066
; REFERENCE/DOCKET NUMBER: 2185-0199P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)205-8000
; TELEFAX: (703)205-8050
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2042 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORGANISM: Daucus carota L.
; INDIVIDUAL ISOLATE: Kuroda Gosun
; FEATURE:
; NAME/KEY: promoter
; LOCATION: 1..2042
; US-08-911-434A-2

Query Match      11.0%; Score 32.8; DB 2; Length 2042;
Best Local Similarity 47.5%; Pred. No. 0.41;
Matches 97; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

Qy 12 CCAGGCGGATGACAGAAAATTCACCTTCCGAGATGCCCTCTAAATCCTGCTGAATTTT 71
Db 197 CCACCGGTAATGGAAGATATGCAAAAAATTCAGCCACAGAGTCTTACTTAACTCT 256
Qy 72 TTACCTTCTCTATGATCGGCAATAGGTTTATTTTCTCAGCTACTTAGTATT 131
Db 257 CACCTTTTGACACACTCTTCTCATATTTTCTTATCTTTTGTCAAAAAATTTGAATA 316
Qy 132 TTGTGGGAGAAGAGGTCACACCCAGACTAATCTTCTCATATGATCTCATGCGAGG 191
Db 317 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 376
Qy 192 CATTGATGATGATGACAGAAAT 215
Db 377 AATAATAATAATAATAATAATAAT 400

RESULT 5
US-09-222-938A-68
; Sequence 68, Application US/09222938A
; Patent No. 6437108
; GENERAL INFORMATION:
; APPLICANT: Youngman, Philip
; APPLICANT: Fritz, Christian
; APPLICANT: Murphy, Christopher
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
; FILE REFERENCE: 07334/060001
; CURRENT APPLICATION NUMBER: US/09/222.938A
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 68
; LENGTH: 800
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:

; NAME/KEY: CDS
; LOCATION: (114)...(743)
; US-09-222-938A-68

Query Match      10.7%; Score 32; DB 4; Length 800;
Best Local Similarity 48.9%; Pred. No. 0.5;
Matches 86; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

Qy 122 ACTTAGTATTTTTGGGAGAAGAGGTCACACCCAGACTAATCTTCTTCTATAATGATCC 181
Db 233 AATTAGTAACCTTACACGACACACACAGAGCCCAAAAACAGTTGACCAAAATTCAGGA 292
Qy 182 TCATGCCAGGCAATTCAGATGATATGACAGAGATCATCTAGGAGGACAAAATGAACCTCAA 241
Db 293 GCAAGTATCAGCTATTCAAGCTGAGCAGTCTAACTTGAAGCTGAAAATGATAGATTACA 352
Qy 242 TGCAGATTCTAGCCAACTGTAAGATGAGACACAGAAAATCGCTGAAAACCTCTATT 297
Db 353 AGCAGAATCTAAGAAACTCGAGGCTGAGATTACAGAACTTTCTAAAAACATTTGTT 408

RESULT 6
US-08-793-410-31
; Sequence 31, Application US/08793410
; Patent No. 5955650
; GENERAL INFORMATION:
; APPLICANT: HITZ, WILLIAM DEAN
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF CANOLA
; TITLE OF INVENTION: AND SOYBEAN PALMIYOYL-ACP THIO-
; TITLE OF INVENTION: ESTERASE GENES AND THEIR USE IN
; TITLE OF INVENTION: THE REGULATION OF FATTY ACID
; TITLE OF INVENTION: CONTENT OF THE OILS OF SOYBEAN
; TITLE OF INVENTION: AND CANOLA PLANTS
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: USA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD VERSION 7.0A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793.410
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10627
; FILING DATE: AUGUST 25, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CHRISTENBURY, LYNNE M.
; REGISTRATION NUMBER: 30,971
; REFERENCE/DOCKET NUMBER: CR-9567-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-5481
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1674 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; US-08-793-410-31

Query Match      10.6%; Score 31.8; DB 2; Length 1674;
Best Local Similarity 51.0%; Pred. No. 0.78;
Matches 75; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
```


RESULT 11

QY 67 ATTTTAACTTCTCTATGATCGCAATAGGGTTTATTTATTTCTCAGCTACTTA 126
||||| || ||| || | || ||| || | ||| |||
Db 345 ATTTTCAATCTTCTTGTGATGTTGAGAGGGTATTTTATTTATGGACCTACTTG 286
||||| || ||| || | || ||| || | ||| |||
QY 127 GTATTTTGGGAGAGAGGGTGACACCCAGACTAATGTTCTTCATAATGATCCTCATG 186
||||| || ||| || | || ||| || | ||| |||
Db 285 ATATTTGATGCTTATGATGTGGAACCTTTTGAATAATGAATATTTCAAATAAATGAAGAGC 226
||||| || ||| || | || ||| || | ||| |||
QY 187 CGAGGCATTCAGATGATAATGGACAGAAATCATCTAGGA 224
|| ||| ||||| ||||| || ||| || |||
Db 225 AAAGAATATCAAGAGATAATGGTCAGAAATTAATTTGA 188
|| ||| ||||| ||||| || ||| || |||

RESULT 13

US-09-180-439-2/c
; Sequence 2, Application US/09180439
; Patent No. 6225532
; GENERAL INFORMATION:
; APPLICANT: Dixon, Mark S
; APPLICANT: Hatzixanthis, Kostas
; APPLICANT: Jones, David A
; APPLICANT: Jones, Jonathan DG
; TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof
; FILE REFERENCE: 620 - 53
; CURRENT APPLICATION NUMBER: US/09/180,439
; CURRENT FILING DATE: 1998-12-06
; EARLIER APPLICATION NUMBER: PCT/GB97/01249
; EARLIER FILING DATE: 1997-05-08
; EARLIER APPLICATION NUMBER: GB 9609681.3
; EARLIER FILING DATE: 1996-05-09
; EARLIER APPLICATION NUMBER: GB 9619924.5
; EARLIER FILING DATE: 1996-09-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 3979
; TYPE: DNA
; ORGANISM: Lycopersicon esculentum
US-09-180-439-2

Query Match 10.0%; Score 30; DB 4; Length 3979;
Best Local Similarity 49.4%; Pred. No. 4.3;
Matches 78; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 67 ATTTTAACTTCTCTATGATCGCAATAGGGTTTATTTATTTCTCAGCTACTTA 126
||||| || ||| || | || ||| || | ||| |||
Db 345 ATTTTCAATCTTCTTGTGATGTTGAGAGGGTATTTTATTTATGGACCTACTTG 286
||||| || ||| || | || ||| || | ||| |||
QY 127 GTATTTTGGGAGAGAGGGTGACACCCAGACTAATGTTCTTCATAATGATCCTCATG 186
||||| || ||| || | || ||| || | ||| |||
Db 285 ATATTTGATGCTTATGATGTGGAACCTTTTGAATAATGAATATTTCAAATAAATGAAGAGC 226
||||| || ||| || | || ||| || | ||| |||
QY 187 CGAGGCATTCAGATGATAATGGACAGAAATCATCTAGGA 224
|| ||| ||||| ||||| || ||| || |||
Db 225 AAAGAATATCAAGAGATAATGGTCAGAAATTAATTTGA 188
|| ||| ||||| ||||| || ||| || |||

RESULT 14

US-09-180-439-7/c
; Sequence 7, Application US/09180439
; Patent No. 6225532
; GENERAL INFORMATION:
; APPLICANT: Dixon, Mark S
; APPLICANT: Hatzixanthis, Kostas
; APPLICANT: Jones, David A
; APPLICANT: Jones, Jonathan DG
; TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof
; FILE REFERENCE: 620 - 53
; CURRENT APPLICATION NUMBER: US/09/180,439
; CURRENT FILING DATE: 1998-12-06
; EARLIER APPLICATION NUMBER: PCT/GB97/01249
; EARLIER FILING DATE: 1997-05-08

; EARLIER APPLICATION NUMBER: GB 9609681.3
; EARLIER FILING DATE: 1996-05-09
; EARLIER APPLICATION NUMBER: GB 9619924.5
; EARLIER FILING DATE: 1996-09-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 4123
; TYPE: DNA
; ORGANISM: Lycopersicon esculentum
US-09-180-439-7

Query Match 10.0%; Score 30; DB 4; Length 4123;
Best Local Similarity 49.4%; Pred. No. 4.4;
Matches 78; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 67 ATTTTAACTTCTCTATGATCGCAATAGGGTTTATTTATTTCTCAGCTACTTA 126
||||| || ||| || | || ||| || | ||| |||
Db 345 ATTTTCAATCTTCTTGTGATGTTGAGAGGGTATTTTATTTATGGACCTACTTG 286
||||| || ||| || | || ||| || | ||| |||
QY 127 GTATTTTGGGAGAGAGGGTGACACCCAGACTAATGTTCTTCATAATGATCCTCATG 186
||||| || ||| || | || ||| || | ||| |||
Db 285 ATATTTGATGCTTATGATGTGGAACCTTTTGAATAATGAATATTTCAAATAAATGAAGAGC 226
||||| || ||| || | || ||| || | ||| |||
QY 187 CGAGGCATTCAGATGATAATGGACAGAAATCATCTAGGA 224
|| ||| ||||| ||||| || ||| || |||
Db 225 AAAGAATATCAAGAGATAATGGTCAGAAATTAATTTGA 188
|| ||| ||||| ||||| || ||| || |||

RESULT 15

US-09-353-585-1/c
; Sequence 1, Application US/09353585
; Patent No. 6287865
; GENERAL INFORMATION:
; APPLICANT: Dixon, Mark S
; Jones, David A
; Jones, Jonathan DG
; TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye PC
; STREET: 8th Floor, 1100 No. 6287865th Glebe Road
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/353,585
; FILING DATE: 15-Jul-1999
; CLASSIFICATION: C12N 15/29, 15/82, A01H 5/00, A01N 65/00, C12Q 1/68
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/930,277
; FILING DATE: 27-OCT-1997
; APPLICATION NUMBER: PCT/GB96/00785
; FILING DATE: 01-APR-1996
; APPLICATION NUMBER: GB 9506658.5
; FILING DATE: 31-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Ms Mary J Wilson
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 620-69
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

LENGTH: 6471 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Tomato
STRAIN: Cf2
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 1754..5012
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 1677..1753
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-353-585-1

Query Match .10.08; Score 30; DB 4; Length 6471;
Best Local Similarity 49.4%; Pred. No. 5.3;
Matches 78; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 67 ATTTTAACTCCCTCTATGGATCGGCAATAGGGTTTATTTTCTCAGCTACTTA 126
Db 1300 ATTTTCAATCTTCTTTGTCATGTTGAGAGGGTATTTTATTTATGGACCTACTTG 1241
QY 127 GTATTTGTTGGGAGAGAGGGTGACACCCAGACTAATGTTCTTCATAATGATCCTCATG 186
Db 1240 ATATTTGATGCTTTATGACGTGGAACCTTTTGAATAATGAATATCAATAATGAAGAGC 1181
QY 187 CGAGGCATTCAGATGATAATGGACAGAAATCATCTAGGA 224
Db 1180 AAAGAATATCAGAGAGATAATGTCAGAAATTAATTTGA 1143

Search completed: June 28, 2003, 00:23:59
Job time : 27.4226 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic : nucleic search, using sw model

Run on: June 27, 2003, 23:29:41 ; Search time 77.4061 Seconds
(without alignments)
5734.045 Million cell updates/sec

Title: US-10-029-359A-1
Perfect score: 299
Sequence: 1 gcaagaggagccacgccc.....tcgctgaaacacctctattag 299

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1055720 seqs, 742224136 residues

Total number of hits satisfying chosen parameters: 2111440

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|-------|---------------|--------|----|---------------------|
| 1 | 299 | 100.0 | 299 | 9 | US-10-029-359A-1 |
| 2 | 236.4 | 79.1 | 1794 | 9 | US-10-235-056-2 |
| 3 | 231 | 77.3 | 1092 | 9 | US-10-235-056-12 |
| 4 | 194 | 64.9 | 1469 | 9 | US-10-235-056-6 |
| 5 | 182.6 | 64.4 | 1092 | 9 | US-10-235-056-14 |
| 6 | 189.2 | 63.3 | 1440 | 9 | US-10-235-056-4 |
| 7 | 187.8 | 62.8 | 1092 | 9 | US-10-235-056-13 |
| 8 | 91 | 30.4 | 1560 | 9 | US-10-029-359A-5 |
| 9 | 66.2 | 22.1 | 555 | 9 | US-10-198-846-13278 |
| 10 | 46.2 | 15.5 | 489 | 9 | US-10-029-359A-3 |
| 11 | 35.6 | 11.9 | 10859 | 9 | US-09-764-891-7408 |
| 12 | 35.2 | 11.8 | 879 | 9 | US-10-198-846-6422 |
| 13 | 33.8 | 11.3 | 618 | 9 | US-10-015-219-96 |
| 14 | 33.8 | 11.3 | 618 | 10 | US-09-777-564-96 |
| 15 | 33.8 | 11.3 | 18692 | 9 | US-10-092-154-1682 |
| 16 | 33.8 | 11.3 | 18692 | 10 | US-09-764-847-1682 |
| 17 | 32.8 | 11.0 | 1919 | 10 | US-09-864-761-10504 |
| 18 | 32.6 | 10.9 | 1975 | 10 | US-09-864-761-15628 |
| 19 | 32.4 | 10.8 | 3996 | 9 | US-10-087-464-42 |

| | | | | | | | |
|---|----|------|------|--------|----|---------------------|--------------------|
| c | 20 | 32.4 | 10.8 | 4095 | 9 | US-09-938-842A-2642 | Sequence 2642, Ap |
| c | 21 | 32 | 10.7 | 408 | 10 | US-09-925-300-351 | Sequence 351, Appl |
| c | 22 | 32 | 10.7 | 800 | 9 | US-10-154-251-68 | Sequence 68, Appl |
| c | 23 | 32 | 10.7 | 800 | 9 | US-10-154-251-69 | Sequence 69, Appl |
| | 24 | 31.8 | 10.6 | 640681 | 10 | US-09-790-988-1 | Sequence 1, Appli |
| | 25 | 31.4 | 10.5 | 470 | 10 | US-09-833-381-1312 | Sequence 1312, Ap |
| c | 26 | 31.4 | 10.5 | 1957 | 10 | US-09-864-761-3787 | Sequence 3787, Ap |
| | 27 | 31.2 | 10.4 | 490 | 9 | US-10-060-036-2349 | Sequence 2349, Ap |
| | 28 | 31.2 | 10.4 | 4023 | 9 | US-09-291-417-17 | Sequence 17, Appli |
| | 29 | 31.2 | 10.4 | 4141 | 9 | US-10-303-683-7 | Sequence 7, Appli |
| | 30 | 31.2 | 10.4 | 4380 | 9 | US-10-303-683-8 | Sequence 8, Appli |
| | 31 | 31 | 10.4 | 6944 | 9 | US-09-790-988-111 | Sequence 111, App |
| c | 32 | 31 | 10.4 | 640681 | 10 | US-09-790-988-1 | Sequence 1, Appli |
| | 33 | 30.8 | 10.3 | 430 | 10 | US-09-960-352-3616 | Sequence 3616, Ap |
| c | 34 | 30.8 | 10.3 | 12744 | 9 | US-09-764-891-8228 | Sequence 8228, Ap |
| c | 35 | 30.6 | 10.2 | 233 | 10 | US-09-983-965-2302 | Sequence 2302, Ap |
| c | 36 | 30.6 | 10.2 | 402850 | 9 | US-09-844-653-5 | Sequence 5, Appli |
| | 37 | 30.4 | 10.2 | 495 | 10 | US-09-728-445-761 | Sequence 761, App |
| c | 38 | 30.4 | 10.2 | 13409 | 9 | US-09-764-891-9601 | Sequence 9601, Ap |
| c | 39 | 30.4 | 10.2 | 13409 | 9 | US-10-205-428-930 | Sequence 930, App |
| c | 40 | 30.2 | 10.1 | 425 | 9 | US-09-918-995-35210 | Sequence 35210, A |
| c | 41 | 30.2 | 10.1 | 668 | 10 | US-09-864-761-32135 | Sequence 32135, A |
| c | 42 | 30.2 | 10.1 | 1233 | 9 | US-10-194-489-2 | Sequence 2, Appli |
| | 43 | 30.2 | 10.1 | 1968 | 10 | US-09-880-107-3711 | Sequence 3711, Ap |
| | 44 | 30 | 10.0 | 313 | 9 | US-10-015-219-550 | Sequence 550, App |
| | 45 | 30 | 10.0 | 313 | 10 | US-09-777-564-550 | Sequence 550, App |

ALIGNMENTS

RESULT 1

US-10-029-359A-1
; Sequence 1, Application US/10029359A
; Publication No. US20030100056A1
; GENERAL INFORMATION:
; APPLICANT: Attersand, Anneli
; TITLE OF INVENTION: Protein Cluster II
; FILE REFERENCE: 10806-156
; CURRENT APPLICATION NUMBER: US/10/029,359A
; CURRENT FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 299
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21)..(299)
; OTHER INFORMATION:
US-10-029-359A-1

Query Match . 100.0%; Score 299; DB 9; Length 299;
Best Local Similarity 100.0%; Pred. No. 3.1e-83;
Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|---|-----|
| Qy | 1 | GCAAGAGGAGCCACGCCGATGACAGAAAATTCACITTCGCGAGATGGCCTCTAAATCCT | 60 |
| Db | 1 | GCAAGAGGAGCCACGCCGATGACAGAAAATTCACITTCGCGAGATGGCCTCTAAATCCT | 60 |
| Qy | 61 | GGCTGAATTTTTTAACTTCTCTATGATCGGCAATAGGCTTTATTTATTTCTCAGC | 120 |
| Db | 61 | GGCTGAATTTTTTAACTTCTCTATGATCGGCAATAGGCTTTATTTATTTCTCAGC | 120 |
| Qy | 121 | TACTTAGTATTTTGGGAGAGGGTGACACCCAGACTAATGTTCTTCTCATATGATC | 180 |
| Db | 121 | TACTTAGTATTTTGGGAGAGGGTGACACCCAGACTAATGTTCTTCTCATATGATC | 180 |
| Qy | 181 | CTCATGGAGGCATTCAGATGATGACAGATCATCTAGGAGGACAAATGAACATCA | 240 |
| Db | 181 | CTCATGGAGGCATTCAGATGATGACAGATCATCTAGGAGGACAAATGAACATCA | 240 |

QY 241 ATGCAGATTCTAGCCACGTAAGATGAGAACACAGAAATCGCTGAAACCTCTATTAG 299
Db 241 ATGCAGATTCTAGCCACGTAAGATGAGAACACAGAAATCGCTGAAACCTCTATTAG 299

RESULT 2

US-10-235-056-2
; Sequence 2, Application US/10235056
; Publication No. US20030059915A1
; GENERAL INFORMATION:
; APPLICANT: Canfield, William M.
; APPLICANT: Cummings, Richard D.
; APPLICANT: Ju, Tongzhong
; TITLE OF INVENTION: CORE 1 3-GALACTOSYL TRANSFERASES
; FILE REFERENCE: 5820.630
; CURRENT APPLICATION NUMBER: US/10/235,056
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US 09/461,321
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1794
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-235-056-2

Query Match 79.1%; Score 236.4; DB 9; Length 1794;
Best Local Similarity 90.8%; Pred. No. 2.5e-63;
Matches 252; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY 22 TGACAGAAATTCACATTCTCGAGATGGCCCTAAATCCTGGCTGAATTTTAAACCTTC 81
Db 40 TTACAGAAATACACTTCGCGAATGGCCCTAAATCCTGGCTGAATTTTAAACCTTC 99
QY 82 TCTATGATCGGCAATAGGCTTTATTTATTTCTCAGCTACTTAGTATTTTGTGGAG 141
Db 100 TCTGTGATCAGCAATAGGATTTCTTTATGTCTCAGCTATTAGTATTTTGTGGAG 159
QY 142 AAGAGGTGACCCAGCACTAATGTTCTTCATATGATCCTCATGCGGCGCATTCAGATG 201
Db 160 AAAAGTTGACCCAGCCCTAATGTTCTTCATATGATCCTCATGCAAGGCATTCAGATG 219
QY 202 ATAATGACAGATCATCTAGGAGCAAAATGAATCTCAATGATGATTTAGCCACGTA 261
Db 220 ATAATGACAGATCATCTAGAGGACAAATGAATGATGATTTAGCCACGTA 279
QY 262 AAGATGAGAACACAGAAATCGCTGAAACCTCTATTAG 299
Db 280 AAGATGAGAACACAGACATTTGCTGAAACCTCTATCAG 317

RESULT 3

US-10-235-056-12
; Sequence 12, Application US/10235056
; Publication No. US20030059915A1
; GENERAL INFORMATION:
; APPLICANT: Canfield, William M.
; APPLICANT: Cummings, Richard D.
; APPLICANT: Ju, Tongzhong
; TITLE OF INVENTION: CORE 1 3-GALACTOSYL TRANSFERASES
; FILE REFERENCE: 5820.630
; CURRENT APPLICATION NUMBER: US/10/235,056
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US 09/461,321
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 1092
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-235-056-12

Query Match 77.3%; Score 231; DB 9; Length 1092;
Best Local Similarity 94.1%; Pred. No. 9.4e-62;
Matches 240; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 45 ATGGCCTCTAAATCCTGGCTGAATTTTAAACCTTCTCTATGATCGGCAATAGGTTT 104
Db 1 ATGGCCTCTAAATCCTGGCTGAATTTTAAACCTTCTCTATGATCGGCAATAGGTTT 60
QY 105 ATTTTATTTCTCAGCTACTTAGTATTTTGTGGGAGAGAGGCTGACCCACGTAAT 164
Db 61 CTTTATGTTCTCAGCTACTTAGTATTTTGTGGGAGAGAGGCTGACCCACGTAAT 120
QY 165 GTTCTTCATATGATCCTCATGCGGCGCATTCAGATGATTAATGACAGAAATCATCTAG 224
Db 121 GTTCTTCATATGATCCTCATGCGGCGCATTCAGATGATTAATGACAGAAATCATCTAG 180
QY 225 GGACAAATGAATCTCAATGCAGATTTCTAGCAAGTAAAGATGAGAACAGAAATCGCT 284
Db 181 GGACAAATGAATCTCAATGCAGATTTCTAGCAAGTAAAGATGAGAACAGAAATCGCT 240
QY 285 GAAACCTCTATTAG 299
Db 241 GAAACCTCTATCAG 255

RESULT 4

US-10-235-056-6
; Sequence 6, Application US/10235056
; Publication No. US20030059915A1
; GENERAL INFORMATION:
; APPLICANT: Canfield, William M.
; APPLICANT: Cummings, Richard D.
; APPLICANT: Ju, Tongzhong
; TITLE OF INVENTION: CORE 1 3-GALACTOSYL TRANSFERASES
; FILE REFERENCE: 5820.630
; CURRENT APPLICATION NUMBER: US/10/235,056
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US 09/461,321
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 1469
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-235-056-6

Query Match 64.9%; Score 194; DB 9; Length 1469;
Best Local Similarity 84.5%; Pred. No. 3.7e-50;
Matches 218; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
QY 42 GAGATGGCCTCTAAATCCTGGCTGAATTTTAAACCTTCTCTATGATCGGCAATAGG 101
Db 177 GAAATGGCCTCTAAATCCTGGCTGAATTTTAAACCTTCTCTATGATCGGCAATAGG 236
QY 102 TTTTATTTTCTCAGCTACTTAGTATTTTGTGGGAGAGAGGCTGACCCACGTA 161
Db 237 TTTTATTTTCTCAGCTACTTAGTATTTTGTGGGAGAGAGGCTGACCCACGTA 296
QY 162 AATGTTCTTCATATGATCCTCATGCGGCGCATTCAGATGATTAATGACAGAAATCATCTA 221
Db 297 AACATGTTTCAATGATCCTCATGCGGCGCATTCAGATGATTAATGACAGAAATCATCTA 356
QY 222 GGAGGCAAAATGAATCTCAATGATGATTTAGGCAAGTAAAGATGAGAACAGAAATC 281
Db 357 AAAGGACAGATGAATCTCAATGATGATTTCCAGGCAACATAAAGATGAGAACAGAAATC 416
QY 282 GCTGAAACCTCTATTAG 299
Db 417 GCTGAGAACCTCTATCAG 434

```
RESULT 5
US-10-235-056-14
; Sequence 14, Application US/10235056
; Publication No. US20030059915A1
; GENERAL INFORMATION:
; APPLICANT: Canfield, William M.
; APPLICANT: Cummings, Richard D.
; APPLICANT: Ju, Tongzhong
; TITLE OF INVENTION: CORE 1 3-GALACTOSYL TRANSFERASES
; FILE REFERENCE: 5820.630
; CURRENT APPLICATION NUMBER: US/10/235,056
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US 09/461,321
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 1092
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-235-056-14

Query Match      64.4%; Score 192.6; DB 9; Length 1092;
Best Local Similarity 84.7%; Pred. No. 8.6e-50;
Matches 216; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 45 ATGGCTCTAAATCCCGCTGAATTTTAAACCTTCTCTATGGATCGGCAATAGGTTT 104
Db 1 ATGGCTCTAAATCTTGGCTGAATTTTAACTTCTCTCTGTGTGATCAGCAATAGGTTT 60

QY 105 ATTTTATTTTCTCAGCTACTTACTTATTTTGGGAGAGAGGGTGACACCCAGCTAAT 164
Db 61 TTTTATGTCTCAACTCTTGAATTTTGTTCGAGAGAGAGGGTGCCATTGACGCTAAC 120

QY 165 GTTCTTCATATGATCTCTCATGCGAGCATTCAGATGATTAATGAGAGATCATCTAGGA 224
Db 121 ATGCTTCATATGACCTCATGCGAGCATTCAGATGATTAATGAGAGATCATCTAGGA 180

QY 225 GGACAAATGAACCTCAATGCGAGCATTCAGATGATTAATGAGAGATCATCTAGGA 284
Db 181 GGACAGATGACCTCAATGCGAGCATTCAGATGATTAATGAGAGATCATCTAGGA 240

QY 285 GAAACCTCTATTAG 299
Db 241 GAGAACCTCTATCAG 255

RESULT 6
US-10-235-056-4
; Sequence 4, Application US/10235056
; Publication No. US20030059915A1
; GENERAL INFORMATION:
; APPLICANT: Canfield, William M.
; APPLICANT: Cummings, Richard D.
; APPLICANT: Ju, Tongzhong
; TITLE OF INVENTION: CORE 1 3-GALACTOSYL TRANSFERASES
; FILE REFERENCE: 5820.630
; CURRENT APPLICATION NUMBER: US/10/235,056
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US 09/461,321
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1440
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-10-235-056-4

Query Match      63.3%; Score 189.2; DB 9; Length 1440;
Best Local Similarity 83.3%; Pred. No. 1.1e-48;
Matches 215; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 42 GAGATGGCTCTAAATCCTGGCTGAATTTTAAACCTTCTCTATGGATCGGCAATAGGG 101
Db 151 GAAATGGCTCTAAATCCTGGCTGAATTTTAACTTCTCTGTGTGATCAGCAATAGGA 210

QY 102 TTTATTTTATTTTCTCAGCTACTTACTTATTTTGGGAGAGAGGGTGACACCCAGCT 161
Db 211 TTTTCTTATGTCTCAACTCTTGAATTTTGTTCGAGAGAGGGTGACGTTACGCT 270

QY 162 AATGTTCTTCATATGATCTCTCATGCGAGCATTCAGATGATTAATGAGAGATCATCTA 221
Db 271 AACATGCTTCATATGATCTCTCATGCGAGCATTCAGACAGATTCAGATATCAGCTC 330

QY 222 GGAGGACAAATGAACCTCAATGCGAGCATTCAGATGATTAATGAGAGATCATCTA 281
Db 331 AAAGGACAGATGAGCTTCAATGCGAGCATTCAGATGATTAATGAGAGATCATCTA 390

QY 282 GCTGAAACCTCTATTAG 299
Db 391 GCTGAGAACCTCTATCAG 408

RESULT 7
US-10-235-056-13
; Sequence 13, Application US/10235056
; Publication No. US20030059915A1
; GENERAL INFORMATION:
; APPLICANT: Canfield, William M.
; APPLICANT: Cummings, Richard D.
; APPLICANT: Ju, Tongzhong
; TITLE OF INVENTION: CORE 1 3-GALACTOSYL TRANSFERASES
; FILE REFERENCE: 5820.630
; CURRENT APPLICATION NUMBER: US/10/235,056
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US 09/461,321
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 1092
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-10-235-056-13

Query Match      62.8%; Score 187.8; DB 9; Length 1092;
Best Local Similarity 83.5%; Pred. No. 2.7e-48;
Matches 213; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 45 ATGGCTCTAAATCCTGGCTGAATTTTAAACCTTCTCTATGGATCGGCAATAGGTTT 104
Db 1 ATGGCTCTAAATCCTTGGCTGAATTTTAACTTCTCTGTGTGATCAGCAATAGGATTT 60

QY 105 ATTTTATTTTCTCAGCTACTTACTTATTTTGGGAGAGAGGGTGACACCCAGCTAAT 164
Db 61 TTCTTATGTCTCAACTCTTGAATTTTGTTCGAGAGAGGGTGACGTTACGCTAAC 120

QY 165 GTTCTTCATATGATCTCTCATGCGAGCATTCAGATGATTAATGAGAGATCATCTAGGA 224
Db 121 ATGCTTCATATGATCTCTCATGCGAGCATTCAGACAGATTCAGATATCAGCTCAA 180

QY 225 GGACAAATGAACCTCAATGCGAGCATTCAGATGATTAATGAGAGATCATCTAGGA 284
Db 181 GGACAGATGAGCTTCAATGCGAGCATTCAGATGATTAATGAGAGATCATCTAGGA 240

QY 285 GAAACCTCTATTAG 299
Db 241 GAGAACCTCTATCAG 255

RESULT 8
US-10-029-359A-5
; Sequence 5, Application US/10029359A
; Publication No. US20030100056A1
; GENERAL INFORMATION:
```



```

US-10-092-154-1682
; Sequence 1682, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092,154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1682
; LENGTH: 18692
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-092-154-1682

```

| | Query Match Best Local Matches | Similarity 56.9% | Pred. No. 11, Conservative 62; | Score 33.8; | DB 9; | Length 18692; | Mismatches 0; | Indels 47; | Gaps 0; |
|----|--------------------------------------|--|--------------------------------------|-------------|-------|---------------|------------------|---------------|------------|
| QY | 139 | GAGAACAGGGTGACACCCAGACTAATGTCTTCATTAATGATCTCATGCGAGGCATTCAG | 198 | | | | | | |
| Db | 3999 | GTGAATTGTATCTGTCGACCCCTTCCAGCCTTAAAGAAAGATCTTCCAGCCGCGCAGCTCAG | 4058 | | | | | | |
| QY | 199 | ATCATTAATGGACAGAAATCATCTAGGAGGACAAATGAAC TTCATTCGAGA | 247 | | | | | | |
| Db | 4059 | ATAATCATGTTCGGAATAGCATATCATGGAATGCAGAAATTCATGGGAGA | 4107 | | | | | | |

Search completed: June 28, 2003, 00:27:34
Job time : 78.4061 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 23:34:12 ; Search time 753.192 Seconds
(without alignments)
6429.241 Million cell updates/sec

Title: US-10-029-359a-1

Perfect score: 299

Sequence: 1 gcaagaggagcagcgccg.....tcgtgaaacacctattag 299

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_hic:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_othr:*

26: em_gss_pro:*

27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-------------|--------------------|
| 1 | 198.2 | 66.3 | 462 | 9 AI236007 | AI236007 EST232569 |
| 2 | 189.2 | 63.3 | 471 | 9 AI059600 | AI059600 UI-R-C1-1 |
| 3 | 182 | 60.9 | 594 | 12 BF471960 | BF471960 UI-M-BH3- |
| 4 | 153.6 | 51.4 | 316 | 12 BE993641 | BE993641 UI-M-BZ1- |
| 5 | 130 | 43.5 | 642 | 9 AJ447748 | AJ447748 AJ447748 |
| 6 | 93.4 | 31.2 | 614 | 9 AJ449471 | AJ449471 AJ449471 |

| | | | | | | |
|----|------|------|------|----|----------|---------------------|
| 7 | 83.4 | 27.9 | 570 | 9 | AL698324 | AL698324 DKF2P686P |
| 8 | 67.6 | 22.6 | 245 | 14 | BQ337716 | BQ337716 CM3-NN002 |
| 9 | 67.2 | 22.5 | 618 | 17 | AZ456919 | AZ456919 IM0260104 |
| 10 | 60.6 | 20.3 | 481 | 12 | BF443807 | BF443807 261354 MA |
| 11 | 60 | 20.1 | 141 | 14 | T10488 | T10488 hbc402 Huma |
| 12 | 40.4 | 13.5 | 701 | 17 | AZ581501 | AZ581501 IM0370M15 |
| 13 | 39.6 | 13.2 | 632 | 12 | BG447948 | BG447948 NF104B09E |
| 14 | 38 | 12.7 | 695 | 13 | BI667399 | BI667399 603292180 |
| 15 | 38 | 12.7 | 696 | 13 | BI102329 | BI102329 602885661 |
| 16 | 37.2 | 12.4 | 541 | 17 | CNS04KUS | AL295309 Tetraodon |
| 17 | 37 | 12.4 | 445 | 12 | BG293789 | BG293789 602390685 |
| 18 | 37 | 12.4 | 712 | 17 | AZ176339 | AZ176339 SP_0141-A |
| 19 | 36.8 | 12.3 | 451 | 17 | AQ337232 | AQ337232 HS-5009-A |
| 20 | 36.8 | 12.3 | 669 | 17 | CNS010M7 | AL099145 Drosophila |
| 21 | 36.6 | 12.2 | 428 | 14 | BQ452715 | BQ452715 sao90e05 |
| 22 | 36.4 | 12.2 | 369 | 13 | BI506540 | BI506540 BB170029A |
| 23 | 36.4 | 12.2 | 709 | 13 | BJ349463 | BJ349463 BJ375529 |
| 24 | 36.4 | 12.2 | 896 | 13 | BJ375529 | BJ375529 BJ375529 |
| 25 | 36.2 | 12.1 | 480 | 10 | AW662066 | AW662066 hll18a03.x |
| 26 | 36.2 | 12.1 | 530 | 9 | AI911553 | AI911553 ty13601.x |
| 27 | 36 | 12.0 | 524 | 17 | B00008 | B00008 CSRL-100b5- |
| 28 | 36 | 12.0 | 678 | 9 | AU237109 | AU237109 AU039635 |
| 29 | 35.8 | 12.0 | 537 | 9 | AU039635 | AU039635 AU039635 |
| 30 | 35.8 | 12.0 | 540 | 9 | AU034415 | AU034415 AU034415 |
| 31 | 35.8 | 12.0 | 769 | 13 | BJ407722 | BJ407722 BJ407722 |
| 32 | 35.8 | 12.0 | 778 | 12 | BG603451 | BG603451 EST502541 |
| 33 | 35.8 | 12.0 | 790 | 13 | BJ403177 | BJ403177 BJ403177 |
| 34 | 35.6 | 11.9 | 522 | 14 | C25513 | C25513 C25513 Dict |
| 35 | 35.6 | 11.9 | 614 | 9 | AU060072 | AU060072 AU060072 |
| 36 | 35.6 | 11.9 | 784 | 17 | BH493958 | BH493958 BOHRK43TF |
| 37 | 35.6 | 11.9 | 836 | 17 | AZ185234 | AZ185234 SP_1004_B |
| 38 | 35.6 | 11.9 | 852 | 14 | BQ171911 | BQ171911 AGENCOURT |
| 39 | 35.6 | 11.9 | 1101 | 17 | CNS00EHW | AL069213 Drosophila |
| 40 | 35.4 | 11.8 | 686 | 13 | BJ361849 | BJ361849 BJ361849 |
| 41 | 35.2 | 11.8 | 368 | 12 | BG526841 | BG526841 40-45 Ste |
| 42 | 35.2 | 11.8 | 578 | 14 | BP007292 | BP007292 BP007292 |
| 43 | 35.2 | 11.8 | 717 | 10 | AV896012 | AV896012 AV896012 |
| 44 | 35 | 11.7 | 904 | 17 | BH161018 | BH161018 ENTRV85TF |
| 45 | 34.8 | 11.6 | 283 | 13 | BMI57689 | BMI57689 NXLV_022_ |

ALIGNMENTS

RESULT 1

AI236007

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AI236007 462 bp mRNA linear EST 31-JAN-1999
EST232569 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
ROVDAL5 3' end, mRNA sequence.

AI236007 1 GI:3829513

EST.

Rattus sp.

Rattus sp.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 462)

Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,

Kerlavage,A.R. and Adams,M.D.

Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat

Gene Index

Unpublished (1998)

Contact: Lee, NH

The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529

Fax: (301)-838-0208

Email: nhlee@tigr.org

Seq primer: M13-21.

Location/Qualifiers

1..462

/organism="Rattus sp."

FEATURES

source

```

/db_xref="ATCC (inhost):2041425"
/db_xref="taxon:10118"
/clone="ROVDA15"
/clone_lib="Normalized rat ovary, Bento Soares"
/clone="Organ: ovary; Vector: pT73Pac; Site_1: EcoRI;
Site_2: NotI"
BASE COUNT 121 a 113 c 117 g 111 t
ORIGIN

Query Match 66.3%; Score 198.2; DB 9; Length 462;
Best Local Similarity 78.9%; Pred. No. 1.9e-43;
Matches 236; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 1 GCAAGAGGAGCCAGCGCGATGACAGAAAATTCACATTTCCGAGATGGCCTCTAAATCCT 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
98 GCGAGAGAGAGCGCGCGTGTGTGAGAAATACATTTCCAGGAATGGCCTCTAAATCCT 157
QY 61 GCGTGAATTTTAACTTCTCTATGATCGGCAATAGGGTTTATTTATTTCTTCAGC 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
158 GCGTGAATTTTAACTTCTCTGAGGATCAGCAATAGGATTTTCTTATGTTCTCAAC 217
QY 121 TACTTAGTATTTTGGGAGAGAGGTGACACCCAGACTAATCTTCTCATATATGATC 180
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
218 TCTTGAATATCTGTGCAAGACAGGCTGACGTTACGCTTAACATGCTTCAATATGATC 277
QY 181 CTCATCGGAGGATTCAGATGATATGGACAGAATCATCTAGGAGGACAAATGAATCTCA 240
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
278 CTCATCGGACATTCAGACGACAGTGGACATAATCACCTCAAGGACAGATGGACTTCA 337
QY 241 ATGCAGATCTAGCCCAACGTAAGATGAGACACAGAAATCGCTGAAACCTCTATTAG 299
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
338 ATGCAGATCTATCCAAACATACAGATGAGAACACAGAGCTTGTGAGAACCTCTATCAG 396

```

```

RESULT 2
AIO59600
LOCUS
DEFINITION
UI-R-CI-lj-g-05-0-UI.s1 UI-R-CI Rattus norvegicus cdna clone
UI-R-CI-lj-g-05-0-UI 3', mRNA sequence.
AIO59600
VERSION
AIO59600.1 GI:3333377
KEYWORDS
EST.
SOURCE
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 471)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL
MEDLINE
9704477
COMMENT
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Oligo-dT track not found, Not I site shown in beginning of sequence
is likely internal to the message. cdna Library Preparation: M.
Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
through Research Genetics This clone is also available through the
I.M.A.G.E. Consortium at LNL (info@image.llnl.gov). IMAGE
ID=1784976
Seq primer: M13 Forward
POLYA-No.
Location/Qualifiers
1..471
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"

```

```

FEATURES
source
1..471

```

```

/clone="UI-R-CI-lj-g-05-0-UI"
/clone_lib="UI-R-CI"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CI
library is a subtracted library derived from the UI-R-C0
library, which is a subtracted library derived from the
UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library
consisted of a mixture of individually tagged normalized
libraries constructed from rat placenta, adult lung,
brain, liver, kidney, heart, spleen, ovary, and muscle.
The UI-R-E1 library consisted of a mixture of
individually tagged normalized libraries constructed from
8, 12 and 18-day embryo. The tag is a string of 3-5
nucleotides present between the Not I site and the
oligo-dT track which allows identification of the library
of origin of a clone within the mixture. The subtracted
library (UI-R-CI) was constructed as follows: PCR
amplified cDNA inserts from UI-R-C0 clones from which 3'
ESTs had been derived was used as a driver in a
hybridization with the UI-R-C0 library in the form of
single-stranded circles. The remaining single-stranded
circles (subtracted library) was purified by
hydroxyapatite column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (Life Technologies) to generate the UI-R-CI
library. This procedure has been previously described
(Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
1996)."
BASE COUNT 122 a 117 c 101 g 131 t
ORIGIN

```

```

Query Match 63.3%; Score 189.2; DB 9; Length 471;
Best Local Similarity 83.3%; Pred. No. 5.3e-41;
Matches 215; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 42 GAGATGGGCTCTAAATCCTGGCTGAATTTTAACTTCTCTATGGATCGCAATAGGG 101
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
169 GAAATGGCTCTAAATCTTGGCTGAATTTTAACTTCTCTATGGATCGCAATAGGA 228
QY 102 TTTATTTATTTTCTCAGCTACTAGTATTTTGGGAGAGAGGGTGACACCCAGACT 161
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
229 TTTTCTTATGTTCTCAACTCTTGAATATTTTGTGCAAGACAGAGCTTTCAGCT 288
QY 162 AATGTTCTCATATGATTCCTCATCGAGGATTCAGATTAATGAGACAGATCATCTA 221
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
289 AACATGCTTCATATGATTCCTCATCGAGACATTCAGACAGAGTGGACATATCACCTC 348
QY 222 GGAGGACAAATGAATTCATTCAGATTCCTAGCCCACTAAAGATGAGAACACAGAAATC 281
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
349 AAAGCACAGATGGACTTCAATGCAGATCTTAGCCCAACATAAAGATGAGAACACAGACGTT 408
QY 282 GCTGAAAACCTCTATTAG 299
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
409 GCTGAGAACCTCTATCAG 426

```

```

RESULT 3
BF471960/c
LOCUS
DEFINITION
UI-M-BH3-awu-c-04-0-UI.r1 NIH_BMAP_M.S4 Mus musculus cdna clone
UI-M-BH3-awu-c-04-0-UI 5', mRNA sequence.
BF471960
ACCESSION
BF471960
VERSION
BF471960.1 GI:11541143
KEYWORDS
EST.
SOURCE
house mouse.
Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 594)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene

```



```

QY 192 CATTGAGATGAATGGACAGATCATCTAGGAGACAAATGAACCTTCAATGAGATTTCT 251
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 129 CATTGAGATGAATGGACAGATCAGCTCAAAAGACAGATGAACCTTCAATGAGATTTCT 188
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 252 ACCCAACGTAAAGATGAGACACAGAAATCGCTGAAACCTCTATTAG `299
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 189 ACCCAACATAAAGATGAGAACATAGACGTTGCTGAGAACCTCTATCAG 236
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
AJ447748
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus
1 (bases 1 to 642)
Buerstedde,J.M.
Gallus gallus bursal lymphocyte EST
Unpublished (2002)
Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinistr. 52, 20251 Hamburg, Germany
Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
Location/Qualifiers
1..642
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone="22cl3rl"
/clone_lib="rikeni"
/cell_type="bursal lymphocyte"
/dev_stage="2-3 weeks old"
/note="CB inbred strain"

BASE COUNT 174 a 135 c 172 g 158 t 3 others
ORIGIN
Query Match 43.5%; Score 130; DB 9; Length 642;
Best Local Similarity 69.7%; Pred. No. 6.5e-25;
Matches 175; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 49 CCTCTAAATCCTCGCTGAATTTTAACTTCCTCTATGGATCGGCAATAGGGTTTATT 108
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 145 CTTCTAAGCTTTGTAGAACCTTCTGACCTTCAGTTTGGATCTGCAATAGGATTTTTC 204
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 109 TATTTTCTCAGCTACTTAGTATTTTGGAGAGAGGGTCACACCCAGACTAATGTTTC 168
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 205 TATGCTATCTCTGTTTGTAGATGATACAGAGAACAGGTTGAGATCAACCTCATATTC 264
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 169 TTCAATGATCCTCATCGAGGCAATTCAGATGATAATGGACAGAAATCATCTAGAGGAC 228
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 265 TTCACATGATCCCCATGGTCAACACTCAGAAGATCTGATAACAATCAGCTGCAAGGAC 324
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 229 AAATGAATTCATGCAATCTAGCCCAACGTAAGATGAGAACACAGAAATCGCTGAAA 288
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 325 AAATGAATTCATGCGGACTCTGGACAGCATAGAGATGAGAAATAGAACATTTGCGATG 384
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 289 ACCTCTATTAG 299
||||| ||||| |||||
Db 385 GACTCTATGAG 395
||||| ||||| |||||

RESULT 6
AJ449471
LOCUS
DEFINITION
ACCESSION

```

```

VERSION
KEYWORDS
SOURCE
ORGANISM
Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus
1 (bases 1 to 614)
Buerstedde,J.M.
Gallus gallus bursal lymphocyte EST
Unpublished (2002)
Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinistr. 52, 20251 Hamburg, Germany
Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
Location/Qualifiers
1..614
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone="22cl3rl"
/clone_lib="rikeni"
/cell_type="bursal lymphocyte"
/dev_stage="2-3 weeks old"
/note="CB inbred strain"

BASE COUNT 185 a 125 c 164 g 140 t
ORIGIN
Query Match 31.2%; Score 93.4; DB 9; Length 614;
Best Local Similarity 69.4%; Pred. No. 5.7e-15;
Matches 127; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 117 CAGCTACTAGTATTTTGTGGAGAGAGGGTGACACCCAGACTAATGTTCTTCAAT 176
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98 CGCGCGCTGAGCATGATACCTAGAGAACAGGTTGAGATCCACCTCATATTCTTCAAT 157
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 177 GATCCTCATGCGAGGATTCAGATGATATGACAGAAATCATCTAGGAGGACAAATGAAC 236
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 158 GATCCCCATGGTCAACACTCAGACATCTGATACATCAGCTGCAAGGACAAATGAAT 217
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 237 TTCAATGAGATTTAGCCACGTTAAAGATGAGACACAGAAATCGCTGAAACCTCTAT 296
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 218 TTCAATGCGGACTCTGGACAGCATAGATGAGAAATAGAACATTCGACATGCACTCTAT 277
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 297 TAG 299
|||
Db 278 GAG 280
|||

RESULT 7
AL698324
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo
1 (bases 1 to 570)
Poustka,A., Wellenreuther,R., Mewes,H.W., Weil,B. and Wiemann,S.).
EST (Poustka,A., Wellenreuther,R., Mewes,H.W., Weil,B. and Wiemann
,S.).
Unpublished (1999)
Contact: Poustka A.J.
Department Lehrach
Max-Planck-Institute for Molecular Genetics
Innestrasse 73, 14195 Berlin, Germany
Tel: +49-30-84131623
Fax: +49-30-84131128
Email: poustka@pimg-berlin-dahlem.mpg.de

AL698324
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo
1 (bases 1 to 570)
Poustka,A., Wellenreuther,R., Mewes,H.W., Weil,B. and Wiemann,S.).
EST (Poustka,A., Wellenreuther,R., Mewes,H.W., Weil,B. and Wiemann
,S.).
Unpublished (1999)
Contact: Poustka A.J.
Department Lehrach
Max-Planck-Institute for Molecular Genetics
Innestrasse 73, 14195 Berlin, Germany
Tel: +49-30-84131623
Fax: +49-30-84131128
Email: poustka@pimg-berlin-dahlem.mpg.de

```

This is the 5' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by DKFZ (German Cancer Research Center,
 Heidelberg/Germany) within the cDNA sequencing consortium of the
 German Genome Project.
 No sl sequence available.
 This clone (DKFZp686P14107) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

```

FEATURES             source
  Location/Qualifiers
    1..570
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone_lib="NN0028"
      /clone_lib="686 (synonym: hlcc3)"
      /tissue_type="human skeletal muscle"
      /dev_stage="adult"
      /lab_host="DH10B"
      /note="vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB;
      cDNA-collection"
BASE COUNT          176 a 116 c 125 g 153 t
ORIGIN
Query Match          27.9%; Score 83.4; DB 9; Length 570;
Best Local Similarity 67.9%; Pred. No. 3e-12;
Matches 131; Conservative 0; Mismatches 61; Indels 1; Gaps 1;
QY 102 TTTATTTTATTTCTCAGCTACTAGTATTGTTGGAGAGAGGCTGACACCCACT 161
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 268 TCTTTTACCTTTTCTCTCTTAATATTCAGTATGGAAGAACAGTCTGACACCT 327
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 162 AATGTTCTCATATGATCTCATCGAGGACATTCAGATGATATGGACAGATCATCTA 221
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 328 GCTAATCTTTGTAACGACTACTCATGTTAGACATTCAGTATGATGGCATACATGTA 387
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 222 GGAGGACAATGAACTTCATGAGATCTAGCCAACTGTAAGATGAGACACAGAATC 281
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 388 -GAGAGATAATGAACTTCAACGGAGGATGCTACCCAAAATCAAGATGGGACACAGATGTT 446
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 282 GCTGAAAACCTCT 294
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 447 ATTGAAAAGCTGT 459

```

```

RESULT 8
BQ337716
LOCUS      BQ337716          245 bp  mRNA  linear  EST 20-MAY-2002
DEFINITION CK3-NN0028-310300-134-e06 NN0028 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BQ337716
VERSION    BQ337716.1  GI:20995424
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 245)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922

```

Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM3&t2=CM3-NN0028-
 310300-134-e06&t3=2000-03-31&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 113
 High quality sequence stop: 127.

```

FEATURES             source
  Location/Qualifiers
    1..245
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone_lib="NN0028"
      /dev_stage="Adult"
      /note="Organ: nervous normal; Vector: puc18; Site_1: SmaI;
      Site_2: SmaI; A mini-library was made by cloning products
      derived from ORESTES PCR (U.S. Letters Patent application
      No. 196,716 - Ludwig Institute for Cancer Research)
      profiles into the pUC 18 vector. Reverse transcription of
      tissue mRNA and cDNA amplification were performed under
      low stringency conditions."
BASE COUNT          74 a 56 c 37 g 78 t
ORIGIN
Query Match          22.6%; Score 67.6; DB 14; Length 245;
Best Local Similarity 77.4%; Pred. No. 5.6e-08;
Matches 82; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
QY 22 TGACAGAAATTCACCTTCGAGATGGCCTCTAAATCCTGGCTGAATTTTAACTTCC 81
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 137 TTACAGAAATACACATTACGGGAAATGGCCTAAATCCTGGCTGAATTTTAACTTCC 196
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 82 TCTATGATCGCAATAGGTTTATTTATTTCTCAGCTACTTAG 127
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 197 TCTGTGATCAGCATAGGATTTCTTTATGTTCCAGCCATTTAG 242

```

```

RESULT 9
A2456919
LOCUS      A2456919          618 bp  DNA  linear  GSS 04-OCT-2000
DEFINITION clone UGCLM0260i04 F, DNA sequence.
ACCESSION  A2456919
VERSION    A2456919.1  GI:10615044
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 618)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0260 row: 1 column: 04
Seq primer: CGTTGTAACACGCGCCAGT
Class: plasmid ends
High quality sequence stop: 618.
FEATURES             source
  Location/Qualifiers
    1..618

```


was directionally synthesized from the Xho I in the vector to the EcoRI site. cDNA was size fractionated to remove sequences <1000 bp in size."

BASE COUNT 57 a 30 c 23 g 31 t
ORIGIN

Query Match 20.1%; Score 60; DB 14; Length 141;
Best Local Similarity 92.4%; Pred. No. 6.3e-06;
Matches 85; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

QY 208 GACAGATCATCTAGGAGGACAAATCAACTTCAATGCAGATTCAGCCACGTAAGATG 267
|||||
Db 1 GACAGATCATCTAGGAGGACAAATCAACTTCAATGCAGATTCAGCCACGTAAGATG 59
|||||

QY 268 AGAACACAGAAATCGCTGAAACCTCTATTAG 299
|||||

Db 60 AGAACACAGACAT-TCTGAAACCTCTATCAG 90
|||||

RESULT 12

AZ581501/c

LOCUS

DEFINITION 701 bp DNA linear GSS 13-DEC-2000
clone UUGC1M0370M15 F, DNA sequence.

ACCESSION

AZ581501

VERSION

GSS.

KEYWORDS

SOURCE

ORGANISM

house mouse.

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 701)

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly

M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.

and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

Plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0370 Row: M Column: 15

Seq primer: CGTTGTAAACGACGCGCCAGT

Class: plasmid ends

High quality sequence stop: 701.

Location/Qualifiers

1..701

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC1M0370M15"

/clone_lib="Mouse, 10kb plasmid UUGC1M library"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"

/note="Vector: pW42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pW42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 240 a 171 c 77 g 213 t
ORIGIN

Query Match 13.5%; Score 40.4; DB 17; Length 701;
Best Local Similarity 51.7%; Pred. No. 1.5;
Matches 92; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 25 CAGAAAATTCACCTTCGAGATGCGCTCTAAATCTGGCTGAATTTTAACTTCCTCT 84
|||||

Db 477 CAGGTTGTTGGTTTCAAAAATAATTAATCTTTCAAGGCAAAAGTATTTTCGATTGTG 418
|||||

QY 85 ATGGATCGCAATAGGTTTATTTTCTCAGCTACTTACTTATTTCTGGGAGAAG 144
|||||

Db 417 ACTGAACACCAATATGAGGTGGTTGTATTAACAGACACATGTTTCTTTAGATGAAG 358
|||||

QY 145 AGGTCGACACCCAGACATAATGTTTCTTAATATGATCTCTATCGAGGCAATTCAGATGA 202
|||||

Db 357 TGGTGGCACTCAGCAGAGATGTCAGACGAGATGAGTGAGGAATTTATTITA 300
|||||

RESULT 13

BG447948

LOCUS

DEFINITION 632 bp mRNA linear EST 16-MAR-2001

clone NF104B09PC 5', mRNA sequence.

ACCESSION BG447948

VERSION BG447948.1

KEYWORDS GI:13366729

SOURCE barrel medic.

ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;

Medicago.

1 (bases 1 to 632)

Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,

Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.

Expressed Sequence Tags from the Samuel Roberts Noble Foundation -

Center for Medicago Genomics Research

Unpublished (2000)

Contact: Dixon RA

Plant Biology Division

The Samuel Roberts Noble Foundation

2510 Sam Noble Parkway, Ardmore, OK 73402, USA

Tel: 580 221 7302

Fax: 580 221 7380

Email: radixon@noble.org

Insert Length: 632 Std Error: 0.00

Plate: 104 row: B column: 09

Seq primer: TCACACAGGAACACGCTATCAC.

Location/Qualifiers

1..632

/organism="Medicago truncatula"

/db_xref="taxon:3880"

/clone="NF104B09PC"

/clone_lib="Elicited cell culture"

/tissue_type="Cell cultures derived from root tissues"

/dev_stage="Cell suspensions were subcultured every 14

days. Cells were induced six days after subculture"

/note="Vector: Lambda Zap; Cells were induced with yeast

cell wall extracts equivalent to 50ug/ml glucose in the

final concentration. Samples were taken at 0.5, 1, 12 and

24 hours after induction. Equal amounts of RNA from each

time point were pooled and used for mRNA isolation."

179 a 108 c 130 g 212 t 3 others

BASE COUNT

ORIGIN

```
Query Match      13.2%; Score 39.6; DB 12; Length 632;
Best Local Similarity 55.6%; Pred. No. 2.4;
Matches 75; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 93 GCATAGGGTTTATTTTCTCAGCTACTTATTTAGTATTTTGGGAGAGGGGTGAC 152
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 151 GCACCAATATATATGTCATGATTTCACCCCACTATATGACATGAGGAAAAGGCC 210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 153 ACCGAGCAATCTTCTCATATGATCTCATGCGGAGGCATTCAGATGATATGACAG 212
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 211 TCACAATTTTTCCTTTAGGAGGAGCCTCACTCAAGATTTTCAGAAGCTGANGTACAT 270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 213 AATCATCTAGGAGGA 227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 271 AACCAACCAAGTGA 285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
BI667399
LOCUS      BI667399
DEFINITION 603292180F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5311723 5',
            mRNA sequence.
ACCESSION  BI667399
VERSION    BI667399.1 GI:15581632
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 695)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-femail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11789 row: o column: 20
            High quality sequence stop: 691.
            Location/Qualifiers
                1..695
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:5311723"
                /clone_lib="NIH_MGC_96"
                /tissue_type="hypothalamus"
                /lab_host="DH10B"
                /notes="Organ: brain; Vector: pBluescriptR (modified
                pBluescript KS+); Site 1: BamHI; Site 2: Sali-XhoI (gtcgag
                ); Oligo-dT primed using primer 5'-TTTTTTTTTTTNN-3',
                size-selected for average insert size 2.3 kb and
                normalized to Rof 5. This is a primary library enriched
                for full-length clones and constructed using the
                Cap-trapper method (Carninci, in preparation). Library
                constructed by M. Brownstein (NIMH/NHGRI, National
                Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT      217 a 119 c 159 g 200 t
ORIGIN

Query Match      12.7%; Score 38; DB 13; Length 695;
Best Local Similarity 51.8%; Pred. No. 6.5;
Matches 86; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 63 CTGAATTTTACCTCTCTCATGATCGGCAATAGGGTTTATTTTCTCAGCTA 122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 514 CTAGTTGTTTAACTCTCTCTTTTACATTTGGCTTTTGTCTTAAATGTTCTCCAAGCTA 573
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Query Match      12.7%; Score 38; DB 13; Length 696;
Best Local Similarity 82.4%; Pred. No. 6.5;
Matches 56; Conservative 0; Mismatches 10; Indels 2; Gaps 1;

QY 234 AACTTCAATCCAGATTCTACCCCAACCTAAGATGAGAACAC--AGAAATCGTGAAC 291
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 AACTTCAATCCAGATTCTACCCCAACCTAAGATGAGAACAC--AGAAATCGTGAAC 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 292 TCTATTAG 299
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 TCTATCAG 68
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: June 28, 2003, 01:00:53
Job time : 757.192 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 22:30:21 ; Search time 1591.73 Seconds
(without alignments)
8940.745 Million cell updates/sec

Title: US-10-029-359A-3
Perfect score: 489
Sequence: 1 catctaaagactgatgaa.....gatatgtactgagcaagaa 489

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|----------|--------------------|
| c 1 | 478 | 97.8 | 153472 | 9 | AC011179 | AC011179 Homo sapi |
| 2 | 478 | 97.8 | 165720 | 9 | AC025524 | AC025524 Homo sapi |
| 3 | 307 | 62.8 | 930 | 9 | HS132443 | AJ132443 Homo sapi |
| 4 | 307 | 62.8 | 1092 | 6 | AX175170 | AX175170 Sequence |
| 5 | 307 | 62.8 | 1092 | 6 | HS243256 | AJ243256 Homo sapi |
| 6 | 307 | 62.8 | 1092 | 9 | HS278960 | AJ278960 Homo sapi |
| 7 | 307 | 62.8 | 1794 | 6 | AX175160 | AX175160 Sequence |
| 8 | 307 | 62.8 | 1794 | 9 | AF155582 | AF155582 Homo sapi |
| 9 | 300.6 | 61.5 | 1092 | 6 | AX175171 | AX175171 Sequence |
| 10 | 300.6 | 61.5 | 1405 | 10 | AF157963 | AF157963 Rattus no |
| 11 | 300.6 | 61.5 | 1440 | 6 | AX175162 | AX175162 Sequence |
| 12 | 292.6 | 59.8 | 1092 | 6 | AX175172 | AX175172 Sequence |
| 13 | 292.6 | 59.8 | 1438 | 10 | AF157962 | AF157962 Mus muscu |
| 14 | 292.6 | 59.8 | 1469 | 6 | AX175164 | AX175164 Sequence |
| 15 | 281 | 57.5 | 1708 | 10 | BC025899 | BC025899 Mus muscu |
| c 16 | 279.8 | 57.2 | 172574 | 2 | AC011121 | AC011121 Homo sapi |
| c 17 | 279.8 | 57.2 | 182823 | 9 | AC016638 | AC016638 Homo sapi |
| c 18 | 279.8 | 57.2 | 217807 | 2 | AC079469 | AC079469 Homo sapi |
| c 19 | 278.2 | 56.9 | 169898 | 2 | AC016585 | AC016585 Homo sapi |
| 20 | 278 | 56.9 | 115954 | 9 | AC005532 | AC005532 Homo sapi |
| 21 | 275 | 56.2 | 146790 | 2 | AC118920 | AC118920 Rattus no |
| c 22 | 275 | 56.2 | 194038 | 2 | AC095494 | AC095494 Rattus no |
| 23 | 270.4 | 55.3 | 1499 | 9 | BC003174 | BC003174 Homo sapi |
| 24 | 244.8 | 50.1 | 123070 | 9 | AC020656 | AC020656 Homo sapi |
| 25 | 244.8 | 50.1 | 170227 | 2 | AC019161 | AC019161 Homo sapi |
| 26 | 244.8 | 50.1 | 307755 | 2 | AC127894 | AC127894 Homo sapi |
| c 27 | 197.8 | 40.4 | 171529 | 2 | AC023021 | AC023021 Homo sapi |
| c 28 | 181 | 37.0 | 183416 | 10 | AL671269 | AL671269 Mouse DNA |
| c 29 | 181 | 37.0 | 229117 | 2 | AL807383 | AL807383 Mus muscu |
| c 30 | 142 | 29.0 | 156337 | 2 | AC009938 | AC009938 Homo sapi |
| c 31 | 116.2 | 23.8 | 167996 | 9 | AC021753 | AC021753 Homo sapi |
| c 32 | 116.2 | 23.8 | 191655 | 9 | AC020661 | AC020661 Homo sapi |
| c 33 | 115.6 | 21.5 | 1170 | 6 | AX175173 | AX175173 Sequence |
| 34 | 105 | 21.5 | 1172 | 3 | AF269063 | AF269063 Caenorhab |
| 35 | 105 | 21.5 | 1172 | 6 | AX175166 | AX175166 Sequence |
| 36 | 105 | 21.5 | 1172 | 6 | AX175176 | AX175176 Sequence |
| 37 | 92.2 | 18.9 | 1167 | 6 | AX175176 | AX175176 Sequence |
| 38 | 92.2 | 18.9 | 2242 | 3 | AY061283 | AY061283 Drosophil |
| 39 | 92.2 | 18.9 | 133090 | 2 | AC017815 | AC017815 Drosophil |
| c 40 | 92.2 | 18.9 | 189904 | 3 | AC093045 | AC093045 Drosophil |
| c 41 | 92.2 | 18.9 | 197901 | 3 | AC009523 | AC009523 Drosophil |
| c 42 | 92.2 | 18.9 | 267547 | 3 | AE003623 | AE003623 Drosophil |
| c 43 | 86.6 | 17.7 | 185872 | 2 | AC068810 | AC068810 Mus muscu |
| 44 | 74.8 | 15.3 | 1104 | 6 | AX175168 | AX175168 Sequence |
| 45 | 74.8 | 15.3 | 1664 | 3 | AY051529 | AY051529 Drosophil |

ALIGNMENTS

RESULT 1
AC011179/c
LOCUS AC011179 153472 bp DNA linear PRI 27-NOV-2001
DEFINITION Homo sapiens, clone RP11-11N24, complete sequence.
ACCESSION AC011179
VERSION AC011179.9 GI:17048255
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 153472)
REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-11N24
JOURNAL Unpublished

TITLE Direct Submission
JOURNAL Submitted (01-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Jan 10, 2002 this sequence version replaced gi:17149773.
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L8338

Center clone name: 705_O_24

FEATURES

```

source      Location/Qualifiers
1..165720
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-705024"
/clone_lib="RPC1-11 Human Male BAC"
repeat_region complement(1..155)
repeat_region /rpt_family="MER4B"
494..523
/rpt_family="AT_rich"
578..599
/rpt_family="AT_rich"
3379..3551
/rpt_family="MLT1J"
3652..3693
/rpt_family="AT_rich"
complement(3993..4312)
/rpt_family="AluY"
5721..5878
/rpt_family="L1MC4a"
5919..6553
/rpt_family="L1MD3"
6564..6592
/rpt_family="AT_rich"
6603..7052
/rpt_family="MLT1C"
7646..7674
/rpt_family="AT_rich"
7978..8107
/rpt_family="MIR"
8524..8589
/rpt_family="L2"
complement(8592..8777)
/rpt_family="MSTB1"
complement(8820..8873)
/rpt_family="MSTB1"
8875..8923
/rpt_family="L2"
complement(10625..11047)
/rpt_family="MLT1C"
12127..12148
/rpt_family="AT_rich"
13527..13813
/rpt_family="AluSx"
13817..13841
/rpt_family="AT_rich"
complement(14152..14273)
/rpt_family="FRAM"
complement(15286..16222)
/rpt_family="L2"
17886..18069
/rpt_family="LTR33"
19251..19312
/rpt_family="(CA)n"
complement(19955..20075)

```

```

repeat_region /rpt_family="L2"
20867..20917
/rpt_family="AT_rich"
21497..21786
/rpt_family="AluSc"
complement(22069..22292)
/rpt_family="L2"
complement(22427..22907)
/rpt_family="L2"
23178..23211
/rpt_family="(TTG)n"
complement(23667..23791)
/rpt_family="L2"
23960..24038
/rpt_family="MLT1K"
24046..24142
/rpt_family="MLT1K"
complement(24144..24231)
/rpt_family="L2"
24669..24693
/rpt_family="(TTTG)n"
complement(26144..26438)
/rpt_family="AluSc"
27613..27662
/rpt_family="(TAGA)n"
complement(28078..28381)
/rpt_family="AluSg"
28947..29054
/rpt_family="L1MB2"
29502..29920
/rpt_family="L1MB2"
31112..31285
/rpt_family="MIR"
complement(31377..31556)
/rpt_family="MIR"
complement(31689..31816)
/rpt_family="MERSA"
31817..31999
/rpt_family="MERSA"
complement(32000..32051)
/rpt_family="MERSA"
complement(32137..32286)
/rpt_family="FRAM"
32589..33275
/rpt_family="MER67C"
33688..33712
/rpt_family="AT_rich"
33827..34528
/rpt_family="MER50"
34559..34759
/rpt_family="MER50-int"
34735..35321
/rpt_family="MER50-int"
35363..35551
/rpt_family="MER67C"
35697..36152
/rpt_family="MER50-int"
complement(36153..36461)
/rpt_family="AluSx"
36462..37932
/rpt_family="MER50-int"
complement(39608..39756)
/rpt_family="L2"
complement(39968..40488)
/rpt_family="L2"
41542..41605

```

Query Match

Best Local Similarity 97.8%; Score 478; DB 9; Length 165720;

Matches 489; Conservative 99.8%; Pred No. 5, 9e-101;

Mismatches 0; Indels 1; Gaps 1;

Qy

1 CATCTAAAGACGTGATGAAGTTGATTCGAATGCTAGTCATCAATAATACCAGAACACA 60
|||||

Db 26488 CATCTAAAAAGACTGATGAAGTTGATTGCAATAGCTAGTCATCAATAATACCAGAACACA 26547
QY 61 GGTGTCACAGCAAACTCTATCAAAAGATGAAATTCCTTCTGCTGATTATGACAGGACCT 120
Db 26548 GGTGTCACAGCAAACTCTATCAAAAGATGAAATTCCTTCTGCTGATTATGACAGGACCT 26607
QY 121 CAAATCTAG-AAAAAGATCAGACGATCAGAGATACATGGGCCAGGCTTGCATATA 179
Db 26608 CAAATCTAGAAAAAAGATCAGACGATCAGAGATACATGGGCCAGGCTTGCATATA 26667
QY 180 ACCGTGTTTATGAGCTCAAAAGAAATAAGACCTTCTTCTACTGTGGGATTACACACAA 239
Db 26668 ACCGTGTTTATGAGCTCAAAAGAAATAAGACCTTCTTCTACTGTGGGATTACACACAA 26727
QY 240 AGAAGACAGAAACCAACTGTCTCTGAAATAGTTAAAGCTTTCTATATGCTCATGACCA 299
Db 26728 AGAAGACAGAAACCAACTGTCTCTGAAATAGTTAAAGCTTTCTATATGCTCATGACCA 26787
QY 300 TTATCTGAATACATGGATTGGTTTCATGAAGCAGATGATGATATATATATATATATC 359
Db 26788 TTATCTGAATACATGGATTGGTTTCATGAAGCAGATGATGATATATATATATATATC 26847
QY 360 ATTGGACAACCTTGAATGGCTTCTCACAACACTATAACCTGATGAATCCACTTACTTTGG 419
Db 26848 ATTGGACAACCTTGAATGGCTTCTCACAACACTATAACCTGATGAATCCACTTACTTTGG 26907
QY 420 GAAAGATTTAAGCACTGCAGAAACAGGACTACATGACTGGAGGAGCAGGATATGACT 479
Db 26908 GAAAGATTTAAGCACTGCAGAAACAGGACTACATGACTGGAGGAGCAGGATATGACT 26967
QY 480 GAGCAAGAA 489
Db 26968 GAGCAAGAA 26977

RESULT 3
HSA132443 930 bp mRNA linear PRI 27-JUL-2001
LOCUS Homo sapiens mRNA for beta-1,3-galactosyltransferase b3gal-T8.
DEFINITION A132443
ACCESSION A132443
VERSION A132443.1 GI:15028815
KEYWORDS b3gal-T8; b3gal-T8 gene; beta-1,3-galactosyltransferase.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 930)
AUTHORS Amado,M., Almeida,R., Schwientek,T. and Clausen,H.
TITLE Identification and characterization of large galactosyltransferase
gene families: galactosyltransferases for all functions
JOURNAL Biochim. Biophys. Acta 1473 (1), 35-53 (1999)
MEDLINE 20047730
PUBMED 10580128

REFERENCE 2 (bases 1 to 930)
AUTHORS Jensen,M.A. and Bennett,E.P.
TITLE Cloning of a new member of the beta-1,3-galactosyltransferase
family, b3gal-T8
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 930)
AUTHORS Jensen,M.A.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-1999) Jensen M.A., School of Dentistry,
University of Copenhagen, Norre Alle 20, 2200 Copenhagen N, DENMARK

FEATURES
source
1..930
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="7p14-p13"
gene 1..930
/gene="b3gal-T8"
CDS 1..930
/gene="b3gal-T8"
/codon_start=1

/product="beta-1,3-galactosyltransferase b3gal-T8"
/protein_id="CAC45046.1"
/db_xref="GI:15028816"
/translation="MASKSWNELTFLCGSAIGLFLCSOLFSLGKVKVDTQPNVLHN
DPARHSDDDNGONHLEQOMNFNDSSQHKDENTDIAENLYOKVILCWVMTGPQNLK
KAKHVKATACRCNKVLFMSSEENKDFPAVLKTEGRDQLYNKTIKAFQVHGHYLE
DADWFLKADDDTYVILDNRLWLLSKSDPEPFIYFGRFPKPVKQVYMGAGVYLSKE
ALRFVDFAFTDKCTHSSIEDLALRCMEIMNVEAGDSNDTIGKETFFHFVPEBHLI
KGLVPRTFWYNNYPPVEVSLLEILLCCQLD"
BASE COUNT 312 a 161 c 192 g 265 t

Query Match 62.88; Score 307; DB 9; Length 930;
Best Local Similarity 78.38; Pred. No. 5.2e-61;
Matches 383; Conservative 0; Mismatches 100; Indels 6; Gaps 1;

QY 1 CATCTAAAAAGACTGATGAAGTTGATTGCAATAGCTAGTCATCAATAATACCAGAACACA 60
Db 172 CATCTAGAACGACAAATGAACCTTCAATGCAAGTTCTAGCAACATAAAGATGAGAACACA 231
QY 61 GGTGTCACCTGACAACTCTATCAAAAGATGAAATTCCTTCTGCTGATATGACAGGACCT 120
Db 232 GACATTTGCTGAAACCTCTATCAGAAAGTTAGAATTCCTTCTGCTGGTTATGACCGCCCT 291
QY 121 CAAAATCTAGAAAAAAGATCAGACGATCAGAGATACATGGGCCAGGCTTGCATATAA 180
Db 292 CAAAACCTAGAGAAAGAGCCAAACACGCTCAAGCTACTTGGGCCAGGCTTGTAAACAA 351
QY 181 GCGTTGTTATGAGCTCAAAAGAAAAATAAAGACTTCTCTACTGTGGGATTAACACCAA 240
Db 352 GTGTTCTTATGAGTTCAAGAGAAATAAAGACTTCCCTGCTGGGACTGAAACCAA 411
QY 241 GAAGACAGAAACCACTGCTCGGAAAAATAGTTAAAGCTTTCTATATGCTCATGACCAT 300
Db 412 GAAGGACAGATCACTATCTGGAACAAATTAAGCTTTTCAGTATCTTCAATGAACAT 471
QY 301 TATCTGGAATACATGAGTTGTTTCATGAAGCAGATGATATATATATATATATACACA 360
Db 472 TATTAGAGATGCTGATGGTTTGAAGCAGATGATGA-----CACGTATGTCATA 525
QY 361 TTGGCAACTTGAATGGCTTCTCACAACATAACCTGATCAATCCACTTACTTTGGG 420
Db 526 CTAGACAATTTGAGGTGGCTTCTTTCAAAATACGACCTGGAAGAACCCATTACTTTGGG 585
QY 421 AAAAGATTTAAGCACTGCGAGAAACAGGACTACATGACTGGAGGAGCAGGATATGACTG 480
Db 586 AGAAGATTTAAGCCCTTATGTAAGCAGGCTACATGAGTGGAGGAGCAGGATATGACTA 645
QY 481 AGCAAGAA 489
Db 646 AGCAAGAA 654

RESULT 4
AX175170 1092 bp DNA linear PAT 03-JUL-2001
LOCUS Sequence 12 from Patent WO0144478.
DEFINITION AX175170
ACCESSION AX175170
VERSION AX175170.1 GI:14598568
KEYWORDS human.
SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1092)
AUTHORS Canfield,W.M., Cummings,R.D. and Ju,T.
TITLE Core lb3-galactosyl transferases and methods of use thereof
JOURNAL Patent: WO 0144478-A 12 21-JUN-2001;
The Board of Regents of The University of Oklahoma (US)
FEATURES
source
1..1092
/organism="Homo sapiens"
/db_xref="taxon:9606"

| | | | | | | | | | |
|-----------------------|---|--|-------|-------|--|--|--|--|--|
| CDS | | | | | | | | | |
| BASE COUNT | 363 a | 195 c | 219 g | 315 t | | | | | |
| ORIGIN | | | | | | | | | |
| Query Match | 62.8%; Score 307; DB 9; Length 1092; | | | | | | | | |
| Best Local Similarity | 78.3%; Pred. No. 5.le-61; | | | | | | | | |
| Matches | 383; Conservative 0; Mismatches 100; Indels 6; Gaps 1; | | | | | | | | |
| QY | 1 | CATCTAAAGACGATGAGCTTGAATGCAAAATGCTAGTCATCATATAATACCAAGACACA | 60 | | | | | | |
| Db | 172 | CATCTAGAAGACAATAAGCTTCAATGCGAGATTTCTAGCCACAATAAAGATGAGAACA | 231 | | | | | | |
| QY | 61 | GGTGTCACTGACAACTCTATCAAAAGATGAAAATCTTTGCTGGATTATGACAGACCT | 120 | | | | | | |
| Db | 232 | GACATGCTGAAACCTCTATCAGAAAGTTAGAAATCTTTGCTGGTTATGACCGCCCT | 291 | | | | | | |
| QY | 121 | CAAAATCTAGAAAAAAGATCAGACGATCAGAGATACATGGGCCAGGTTGCAATAAA | 180 | | | | | | |
| Db | 292 | CAAAACCTTAGAAAAAAGGCCAAACAGCTCAAAAGCTACTTTGGGCCAGCTTCTTAACAA | 351 | | | | | | |
| QY | 181 | CGGTGTTTATGAGCTCAAAAGAAATTAAGACTTCTCTACTGTGGGATTTACACACCA | 240 | | | | | | |
| Db | 352 | GTGTGTTTATGAGTTTCAAGAAAAATTAAGACTTCCCTGCTGTGGGACTGAAACCCAA | 411 | | | | | | |
| QY | 241 | GAAGACAAAACCAACTGCTCCTGGAAAAATAGTTAAAGCTTTTCTATATGCTCATGACC | 300 | | | | | | |
| Db | 412 | GAAGCAGAGATCAACTATATCTGGAACAACTTAAGACTTTTCAGTATGTTTCATGA | 471 | | | | | | |
| QY | 301 | TATCTGGAATACATGGATTGGTTTCATGAAACGAGATGATATGATATATATATACAC | 360 | | | | | | |
| Db | 472 | TATTTAGAAGATGCTGATTGGTTTTTTTGAAGCAGATGATGA-----CACGTATG | 525 | | | | | | |
| QY | 361 | TTGGACACTTGAATGCCTCTCACAACACTATAACCTCATGATCAATCCACTTACTTTGG | 420 | | | | | | |
| Db | 526 | CTAGACAAATTTGAGTGGCTTCTTTCAAAATACGACCTGAAAGACCCATTTACTTTGG | 585 | | | | | | |
| QY | 421 | AAAGATTTAAGCACTGCGACAAAACAGACTACATGACTGGAGAGCAGGATATGTACTG | 480 | | | | | | |
| Db | 586 | AGAAGATTTAAGCCCTTATGTAACAGCGGCTACATGATGAGTGGAGCAGGATGTACTA | 645 | | | | | | |
| QY | 481 | AGCAAAAGAA 489 | | | | | | | |
| Db | 646 | AGCAAAAGAA 654 | | | | | | | |
| RESULT 6 | | | | | | | | | |
| LOCUS | HSA278960 | | | | | | | | |
| DEFINITION | Homo sapiens mRNA for beta 1,3-galactosyltransferase (core-1-beta | | | | | | | | |
| ACCESSION | AJ278960 | | | | | | | | |
| VERSION | AJ278960.1 GI:17384685 | | | | | | | | |
| KEYWORDS | beta 1,3-galactosyltransferase; core-1-beta | | | | | | | | |
| SOURCE | human. | | | | | | | | |
| ORGANISM | Homo sapiens | | | | | | | | |
| REFERENCE | 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | | | | | |
| AUTHORS | Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. | | | | | | | | |
| Bennett,E.P. | | | | | | | | | |

| | | | | | | | | | |
|-----------------------|---|---|-------|-------|--|--|--|--|--|
| CDS | | | | | | | | | |
| BASE COUNT | 363 a | 195 c | 219 g | 315 t | | | | | |
| ORIGIN | | | | | | | | | |
| Query Match | 62.8%; Score 307; DB 6; Length 1092; | | | | | | | | |
| Best Local Similarity | 78.3%; Pred. No. 5.le-61; | | | | | | | | |
| Matches | 383; Conservative 0; Mismatches 100; Indels 6; Gaps 1; | | | | | | | | |
| QY | 1 | CATCTAAAGACGATGAGCTTGAATGCAAAATGCTAGTCATCATATAATACCAAGACACA | 60 | | | | | | |
| Db | 172 | CATCTAGAAGACAATAAGCTTCAATGCGAGATTTCTAGCCACAATAAAGATGAGAACA | 231 | | | | | | |
| QY | 61 | GGTGTCACTGACAACTCTATCAAAAGATGAAAATCTTTGCTGGATTATGACAGACCT | 120 | | | | | | |
| Db | 232 | GACATGCTGAAACCTCTATCAGAAAATTAAGACTTCCCTGCTGTGGGACTGAAACCCAA | 291 | | | | | | |
| QY | 121 | CAAAATCTAGAAAAAAGATCAGACGATCAGAGATACATGGGCCAGGTTGCAATAAA | 180 | | | | | | |
| Db | 292 | CAAAACCTTAGAAAAAAGGCCAAACAGCTCAAAAGCTACTTTGGGCCAGCGTTGTACAA | 351 | | | | | | |
| QY | 181 | CGGTGTTTATGAGCTCAAAAGAAATTAAGACTTCTCTACTGTGGGATTACACACAA | 240 | | | | | | |
| Db | 352 | GTGTGTTTATGAGTTTCAAGAAAAATTAAGACTTCCCTGCTGTGGGACTGAAACCCAA | 411 | | | | | | |
| QY | 241 | GAAGACAAAACCAACTGCTCTCGAAAAATAGTTAAAGCTTTTATATGCTCATGACC | 300 | | | | | | |
| Db | 412 | GAAGCAGAGATCAACTATCTAGGAAAACAAATTAAGCTTTTTCAGTATGTTTCATGA | 471 | | | | | | |
| QY | 301 | TATCTGGAATACATGGATTGTTTCATGAAAGCAGATGATATATGATATATATATACA | 360 | | | | | | |
| Db | 472 | TATTTAGAAGATGCTGATTGGTTTTTTGAAGCAGATGATGA-----CACGTATG | 525 | | | | | | |
| QY | 361 | TTGGACACTTGAATGGCTTCTCACAACACTATAACCTGATGAATCCACTTACTTTGG | 420 | | | | | | |
| Db | 526 | CTAGACAAATTTGAGTGGCTTCTTTCAAAATACGACCTTGAAGACCCATTTACTTTGG | 585 | | | | | | |
| QY | 421 | AAAGATTTAAGCACTGCGACAAAACAGACTACATGACTGGAGAGCAGGATATGTACTG | 480 | | | | | | |
| Db | 586 | AGAAGATTTAAGCCCTTATGTAACAGCGGCTACATGATGAGTGGAGGAGCAGGATGTACTA | 645 | | | | | | |
| QY | 481 | AGCAAAAGAA 489 | | | | | | | |
| Db | 646 | AGCAAAAGAA 654 | | | | | | | |
| RESULT 5 | | | | | | | | | |
| LOCUS | HSA243256 | | | | | | | | |
| DEFINITION | Homo sapiens mRNA for beta-1,3-galactosyltransferase. | | | | | | | | |
| ACCESSION | AJ243256 | | | | | | | | |
| VERSION | AJ243256.1 GI:18073134 | | | | | | | | |
| KEYWORDS | b3Gal-T gene; beta-1,3-galactosyltransferase. | | | | | | | | |
| SOURCE | human. | | | | | | | | |
| ORGANISM | Homo sapiens | | | | | | | | |
| REFERENCE | 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | | | | | |
| AUTHORS | Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. | | | | | | | | |
| JOURNAL | Jensen,M.P.A. | | | | | | | | |
| TITLE | Cloning and expression of a novel beta-1,3-Galactosyltransferase | | | | | | | | |
| REFERENCE | 2 (bases 1 to 1092) | | | | | | | | |
| AUTHORS | Jensen,M.P.A. | | | | | | | | |
| TITLE | Direct Submission | | | | | | | | |
| JOURNAL | Submitted (23-JUN-1999) Jensen M.P.A., School of Dentistry, | | | | | | | | |
| UNIVERSITY | University of Copenhagen, Noerre Allé 20, DK-2200 Copenhagen N, | | | | | | | | |
| DENMARK | | | | | | | | | |
| FEATURES | Location/Qualifiers | | | | | | | | |
| source | 1..1092 | | | | | | | | |
| | /organism="Homo sapiens" | | | | | | | | |
| | /db_xref="taxon:9606" | | | | | | | | |
| | /chromosome="7" | | | | | | | | |
| | /map="p14-p13" | | | | | | | | |
| gene | 1..1092 | | | | | | | | |
| | /gene="b3Gal-T" | | | | | | | | |

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1092)
AUTHORS Bennett,E.P.
TITLE Direct Submission
JOURNAL Submitted (06-SEP-2000) Bennett E.P., Histo. lab., 24 5.33, School
of Dentistry, Noerre Alle' 20, 2200N, Copenhagen, DENMARK

FEATURES

Location/Qualifiers

1..1092
/organism="Homo sapiens"
/db_xref="taxon:9606"

gene

1..1092
/gene="core-1-beta 1,3-galactosyltransferase"

CDS

1..1092
/function="core-1 synthase"

/codon_start=1
/evidence=experimental

/product="beta 1,3-galactosyltransferase"

/protein_id="CAC82373.1"

/db_xref="GI:17384686"

/translation="MASKSWLNFLFLCGSAIGFLCSQLFSILIGKVDTPQNVLHN
DPHARHSDNONGHLEGMNFNADSSQHKDENTDIAENLYQKVRILCWMTGPONLEK
KAKHVKAQWQRCNKVLFMSSEENKDPFVGLKTEGRDQLYWTKIKAFQVHHEYLE
DADWFLKADDDTYVILNRLWLLSKYDPEEPIYGRFRKPYVKQYMSGGAGYVLSKE
ALKRFVAFKTDKCTHSSSIDLALGRCEIMNVEAGDSRDTIGKETPHFVPEHHLI
KGYLPTFRFWYNNYYPVEGPGCCSLAVSFHYVDSTMYELEYLVHLRPGYLYR
YQPTLPERILKEISOANKNEDTKVLGNP"

BASE COUNT 363 a 195 c 219 g 315 t

ORIGIN

Query Match 62.8%; Score 307; DB 9; Length 1092;

Best Local Similarity 78.3%; Pred. No. 5.1e-61;

Matches 383; Conservative 0; Mismatches 100; Indels 6; Gaps 1;

1 CATCTAAAAAGACTGATGAAGTTGATTGCAAAATGCTAGTCATCAATAATACCAGAACACA 60

172 CATCTAGAAGGACAAATGAATTCATCGACATTCAGCCACATATAAGATGAGAACACA 231

61 GGTGTCACAGCAAACTCTATCAAAAGATGAAAATCTTTGCTGATTATGACAGACCT 120

232 GACATGCTGAAAACCTCTATCAGAAAAGTTAGAATTCCTTGTGGTTTATGACCGCCCT 291

121 CAAATCTAGAAAAAGATCAGCGCATCAGAGATACATGGGCCCGGTTGCAATAAA 180

292 CAAAACCTTAGAAAAAGGCCAACACGCTCAAAGCTACTTGGGCCCGGTTGTAAACAA 351

181 GCGTTGTTTATGAGCTCAAAAGAAATAAAGACTTCTCTACTGTGGGATTACACACCAA 240

352 GTGTTGTTTATGAGTTAGAGAAATAAAGACTTCTCTGCTGTGGGACTGAAAACCAA 411

241 GAAGACAGAAACCACTGCTCTGGAATAATAGTTAAAGCTTTTCTATATGCTCATGACCAT 300

412 GAAGCAGAGATCAACTACTTCTGGAATAATAGTTAAAGCTTTTCTATATGCTCATGACCAT 471

301 TATCTGGAATACATGAGTTGTTCTATGAAGCAGATGATATATATATATATATATATAT 360

472 TATTAGAAAGATGCTGATTGTTTGTGAAGCAGATGATGA-----CACGTTATGTCATA 525

361 TTGGACAACCTTGAATGCTCTCAACAACTATACCTGATGATGATGATGATGATGATGATG 420

526 CTAGACAATTTGAGTGCTCTCTCAAAATACGACCTTCAAGAACCCATTTTACCTTTGGG 585

421 AAAAGATTAAAGCACTGCGAGAAACAGCACTACATGACTGGAGGAGCAGGATATGACTG 480

586 AGAAGATTAAAGCCCTTATGTAAGCAGGCGCTACATGAGTGGAGGAGCAGGATATGACTA 645

481 AGCAAGAA 489

646 AGCAAGAA 654

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

Source

CDS

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Score

DB

Length

Indels

Gaps

1

CATCTAAAAAGACTGATGAAGTTGATTGCAAAATGCTAGTCATCAATAATACCAGAACACA 60

234 CATCTAGAAGGACAAATGAATTCATCGACATTCAGCCACATATAAGATGAGAACACA 293

61 GGTGTCACAGCAAACTCTATCAAAAGATGAAAATCTTTGCTGATTATGACAGACCT 120

294 GACATGCTGAAAACCTCTATCAGAAAAGTTAGAATTCCTTGTGGTTTATGACCGCCCT 353

121 CAAATCTAGAAAAAGATCAGCGCATCAGAGATACATGGGCCCGGTTGCAATAAA 180

354 CAAAACCTTAGAAAAAGGCCAACACGCTCAAAGCTACTTGGGCCCGGTTGTAAACAA 413

181 GCGTTGTTTATGAGCTCAAAAGAAATAAAGACTTCTCTACTGTGGGATTACACACCAA 240

414 GTGTTGTTTATGAGTTAGAGAAATAAAGACTTCTCTGCTGTGGGACTGAAAACCAA 473

241 GAAGACAGAAACCACTGCTCTGGAATAATAGTTAAAGCTTTTCTATATGCTCATGACCAT 300

474 GAAGCAGAGATCAACTACTTCTGGAATAATAGTTAAAGCTTTTCTATATGCTCATGACCAT 533

301 TATCTGGAATACATGAGTTGTTCTATGAAGCAGATGATATATATATATATATATATAT 360

534 TATTAGAAAGATGCTGATTGTTTGTGAAGCAGATGATGA-----CACGTTATGTCATA 587

361 TTGGACAACCTTGAATGCTCTCAACAACTATACCTGATGATGATGATGATGATGATGATG 420

588 CTAGACAATTTGAGTGCTCTCTCAAAATACGACCTTCAAGAACCCATTTTACCTTTGGG 647

421 AAAAGATTAAAGCACTGCGAGAAACAGCACTACATGACTGGAGGAGCAGGATATGACTG 480

648 AGAAGATTAAAGCCCTTATGTAAGCAGGCGCTACATGAGTGGAGGAGCAGGATATGACTA 707

481 AGCAAGAA 489

708 AGCAAGAA 716


```
QY 181 GCCTGCTTTATGAGCTCAAAAGAAATAAAGACTTCTCTACTGTGGGATTACACACAAA 240
Db 352 GTGCTTTATGAGTTTCAAGAGAAATAAAGACTTCTCTACTGTGGGTTGAAACAAA 411
QY 241 GAAGACAGAAACCAACTGCTCCCTGGAATAAGTAAAGCTTTCTATATCTCATGACCAT 300
Db 412 GAAGCAGAGCAAGCACTGCTACGAAACAAATTAAGCGTTTCAGTACGATACATGACCAT 471
QY 301 TATCTGGAATACATGATGGTGGTTCATGAAAGCAGATGATATATATATATATATACCA 360
Db 472 TACTTAGAAGATGCTGACTGGTTTATGAAAGCAGACGATGA-----CACCTATGTCATA 525
QY 361 TTGGCAACTTGAATGGCTTCTCAAAACTATAACCCCTGATGAATCCACTTACTTTGGG 420
Db 526 CTGGACAATCTGAGATGGCTTCTATCAAAAGTATAAAGCTTCAACAGCCCATTTACTTTGGG 585
QY 421 ARAAGATTAAAGCACTGCAGAAACAGGACTACATGACTGGAGCAGAGATGATGACG 480
Db 586 AGAAGATTAAAGCCCTATGTGAAGCAGGATACATGAGTGGAGGACGAGATATGTCCTA 645
QY 481 AGCAAGAA 489
Db 646 AGCAAGAA 654

RESULT 10
AF157963
LOCUS
DEFINITION
Rattus norvegicus corel UDP-galactose:N-acetylglactosamine-alpha-R
beta 1,3-galactosyltransferase (Cigalt1) mRNA, complete cds.
AF157963
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 1405)
Ju,T., Brewer,K., D'Souza,A., Cummings,R.D. and Canfield,W.M.
Cloning and Expression of Human Core 1 beta
1,3-Galactosyltransferase
J. Biol. Chem. 277 (1), 178-186 (2002)
11677243
2 (bases 1 to 1405)
Ju,T.Z., Cummings,R.D. and Canfield,W.M.
Cloning and characterization of corel
UDP-galactose:N-acetylglactosamine-alpha-R beta
1,3-galactosyltransferase (Corel beta3-Gal-T) cDNA sequence
Unpublished
3 (bases 1 to 1405)
Ju,T.Z., Cummings,R.D. and Canfield,W.M.
Direct Submission
Submitted (10-JUN-1999) Department of Medicine, The University of
Oklahoma, Health Science Center, 975 N.E. 10th Street, Biomedical
Research Center, Room 411, Oklahoma City, OK 73104, USA
Location/Qualifiers
1. 1405
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
1. 1405
/gene="Cigalt1"
154. 1245
/gene="Cigalt1"
/EC_number="2.4.1.122"
/function="responsible for the synthesis of corel
structure of O-glycans"
/notes="corel beta3-Gal-T"
/codon_start=1
/product="corel
UDP-galactose:N-acetylglactosamine-alpha-R beta
1,3-galactosyltransferase"
/protein_id="AAF81983.1"
```

```
/db_xref="GI:8927166"
/translation="MASKSWLNFLFLCGSAIGFLCSQLLNILLQEQADVPQMLHN
DPHARISDSGHNHLKQMDNFNADSSOHKNDVAENLYQKVLKWMYSPONLEK
KAKHKVATWAORCNKVLFMSSSENKDFPTVGLTEKGRBOLYKTKAFOYVHDVLE
DADWENKADDDTVYILDNLRLWLLSKYNPEQYFGRFRPYKQGYMSGAGYVLSKE
ALURVEDAKTEKCTHSSSDIALGRCHMEIINKVEAGDSRDPTGKTFIPFVEHHLI
KGYLPKTFWNNYNNYPVPGCGCSDIAVFSHYVDSTTMYELEYLVYHLRPYGYLYR
YQPALPENILKEINQVKKEDTKIKLGNP"
misc_feature
166. 249
/gene="Cigalt1"
/notes="Region: non-cleavable transmembrane signal"
/evidence=not_experimental
BASE COUNT 445 a 288 c 300 g 372 t
ORIGIN
Query Match 61.5%; Score 300.6; DB 10; Length 1405;
Best Local Similarity 77.5%; Pred. No. 1.5e-59;
Matches 379; Conservative 0; Mismatches 104; Indels 6; Gaps 1;
QY 1 CATCTAAAAGACTGATGAAGTTGATTGCAAAATGCTAGTCTCATCAATAATACCAACACA 60
Db 325 CACCTCAAAGGACAGATGGACTTCAATGCAGATCTAGCCCAACATAAAGATGAGAACACA 384
QY 61 GGTGTCACAGCAAACTCTATCAAAAGATGAAAATTTCTTTCTGGGATTATGACAGGACCT 120
Db 385 GACGTTCTGAGAACCTCTATCAGAAAGTTAAAGTTCTTTTGGTTTATGACAAGCCCT 444
QY 121 CAAAATCTAGAAAAAAGATCAGACGCATCAGAGATACATGGCCGACGGTTTCAATAAA 180
Db 445 CAAAATCTAGAAAAAAGGCCAACACAGTCAAGTACATGGGCCCGCCGCTTAAATAAA 504
QY 181 GCGTTGTTATGAGCTCAAAAGAAATAAAGACTTCTCTACTGTGGGATTACACACCAA 240
Db 505 GTCTTATTATGAGTTTCAGAGAATAAAGACTTCCCTACTGTGGGCTTGAACACCAA 564
QY 241 GAAGACAGAAACCAACTGCTCTGGAATAATAGTTTAAAGCTTTTCTATATGCTCATGACCAT 300
Db 565 GAAGGAGGAGAGCAACTGTACTGGAAAAACAATTTAAAGCGTTTCAGTACGTACATGACCAT 624
QY 301 TATCTGGAATACATGAGTGGTTTCATGAAGCAGATGATCATATATATATATATATACACA 360
Db 625 TACTTAGAAGATGCTGACTGGTTTATGAAGAGCAGACGATGA-----CACCTATGTCATA 678
QY 361 TTGGCAACTTGAATGGCTTCTCAAAACATATAACCCCTCATGAATCCACTTACTTTGGG 420
Db 679 CTGGACAATCTGAGATGGCTTCTCTCAAGTATATCAACCTGAAACAGCCATTTACTTTGGG 738
QY 421 AAAAGATTTAAGCACTGCAGAAACAGGACTACATGACTGGAGGACGAGGATATGTAAGT 480
Db 739 AGAAGATTTAAGCCCTATGTGAAGCAGGATACATGAGTGGAGGACGAGGATATGTCCTA 798
QY 481 AGCAAGAA 489
Db 799 AGCAAGAA 807

RESULT 11
AX175162
LOCUS
DEFINITION
Sequence 4 from Patent WO0144478.
AX175162
ACCESSION
AX175162.1 GI:14598564
VERSION
KEYWORDS
SOURCE
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 1440)
Canfield,W.M., Cummings,R.D. and Ju,T.
Core lb3-galactosyl transferases and methods of use thereof
Patent: WO 0144478-A 4 21-JUN-2001;
The Board of Regents of The University of Oklahoma (US)
```

| FEATURES | | Location/Qualifiers | | Matches 374; Conservative 0; Mismatches 109; Indels 6; Gaps 1; | |
|--|-----|--|-----|--|---|
| source | | 1. .1440 | | | |
| | | /organism="Rattus norvegicus" | | | |
| | | /db_xref="taxon:10116" | | | |
| BASE COUNT | | 450 a 302 c 311 g 377 t | | | |
| ORIGIN | | | | | |
| Query Match | | 61.5%; Score 300.6; DB 6; Length 1440; | | | |
| Best Local Similarity | | 77.5%; Pred. No. 1.5e-59; | | | |
| Matches 379; Conservative 0; Mismatches 104; Indels 6; Gaps 1; | | | | | |
| Qy | 1 | CATCTAAAAAGACTGATGAAGTTGATTGCAATGCTAGTCATCATATAAATACCAAGACACA | 60 | 1 | CATCTAAAAAGACTGATGAAGTTGATTGCAATGCTAGTCATCATATAAATACCAAGACACA |
| Db | 325 | CACCTCAAGGACGAGATGGACTTCAATGCAGATTCAGCAACATATAAAGATGAGAACACA | 384 | 172 | CACCTCAAGGACGAGATGAACCTTCAATGCAGATTCAGCAACATATAAAGATGAGAACACA |
| Qy | 61 | GGTGTCACTGACAACTCTATCAAAAGATCAAAATCTTTGCTGGGATTATGACAGGACCT | 120 | 61 | GGTGTCACTGACAACTCTATCAAAAGATCAAAATCTTTGCTGGGATTATGACAGGACCT |
| Db | 385 | GACGTTCTGAGAACCTCTATCAGAAAGTTAAAGTTCTTTGTTGGGTTATGACAAGCCCT | 444 | 232 | GACGTTCTGAGAACCTCTATCAGAAAGTTAAAGTTCTTTGTTGGGTTATGACAAGTCCT |
| Qy | 121 | CAAAATCTAGAAAAAGATCAGAGCATCAGAGATACATGGGCCAGGTTGCAATATAA | 180 | 121 | CAAAATCTAGAAAAAGATCAGAGCATCAGAGATACATGGGCCAGGTTGCAATATAA |
| Db | 292 | CAAAATCTAGAAAAAGGCCAAACATGTCAAAGTACGTGGGCCAGCGTTGTGTAATAA | 351 | 292 | CAAAATCTAGAAAAAGGCCAAACATGTCAAAGTACGTGGGCCAGCGTTGTGTAATAA |
| Qy | 181 | GCCTGTTTATGAGCTCAAAAGAAATAAAGACTTCTCTACTGTGGGATTACACACAAA | 240 | 181 | GCCTGTTTATGAGCTCAAAAGAAATAAAGACTTCTCTACTGTGGGATTACACACAAA |
| Db | 352 | GTGTTTATGAGTTGCGGAAGAAATCAAGACTTCCTCTACTGTGGGATTGAAACCAAA | 411 | 352 | GTGTTTATGAGTTGCGGAAGAAATCAAGACTTCCTCTACTGTGGGATTGAAACCAAA |
| Qy | 241 | GAAGACAGAAACCAACTGCTCTGGAATAATAGTTAAAGCTTTTCTATATGCTCATGACCAT | 300 | 241 | GAAGACAGAAACCAACTGCTCTGGAATAATAGTTAAAGCTTTTCTATATGCTCATGACCAT |
| Db | 412 | GAGGACAGAGCAACTATATTGGAACAATTAAGCTTTCCAGTATGTACATGACCAT | 471 | 412 | GAGGACAGAGCAACTATATTGGAACAATTAAGCTTTCCAGTATGTACATGACCAT |
| Qy | 301 | TATCTGGAATACATGGATTGGTTTCATGAAAGCAGATGATATATATATATATATACCA | 360 | 301 | TATCTGGAATACATGGATTGGTTTCATGAAAGCAGATGATATATATATATATATACCA |
| Db | 472 | TATTTAGAGATGCTGACTGGTTTATGAAAGCAGATGACGA-----CACATACGTCTT | 525 | 472 | TATTTAGAGATGCTGACTGGTTTATGAAAGCAGATGACGA-----CACATACGTCTT |
| Qy | 361 | TTGACACAACTGAAATGGCTTCTCACAACATATAACCTGATGAATCCACTTACTTTGGG | 420 | 361 | TTGACACAACTGAAATGGCTTCTCACAACATATAACCTGATGAATCCACTTACTTTGGG |
| Db | 526 | GTGACAACTGAGATGGCTTCTATCAAGTATAACCTGAACACCCATTACTTTTGGG | 585 | 526 | GTGACAACTGAGATGGCTTCTATCAAGTATAACCTGAACACCCATTACTTTTGGG |
| Qy | 421 | AAAAGATTAAAGCACTGCAGAACAGGACTACATGACTGGAGGAGGAGGATGTACTG | 480 | 421 | AAAAGATTAAAGCACTGCAGAACAGGACTACATGACTGGAGGAGGAGGATGTACTG |
| Db | 586 | CGAAGATTAAAGCCCTATGTGAAGCAGGATACATGAGCGGAGGAGGCGGCTATGTCTTA | 645 | 586 | CGAAGATTAAAGCCCTATGTGAAGCAGGATACATGAGCGGAGGAGGCGGCTATGTCTTA |
| Qy | 481 | AGCAAAGAA 489 | | 481 | AGCAAAGAA 489 |
| Db | 646 | AGCAAAGAA 654 | | 646 | AGCAAAGAA 654 |
| RESULT 13 | | 1438 bp mRNA linear ROD 03-JAN-2002 | | | |
| AF157962 | | Mus musculus core UDP-galactose:N-acetylgalactosamine-alpha-R beta | | | |
| LOCUS | | 1.3-galactosyltransferase (Clgal1) mRNA, complete cds. | | | |
| DEFINITION | | AF157962 | | | |
| ACCESSION | | AF157962.1 GI:8927163 | | | |
| VERSION | | | | | |
| KEYWORDS | | Mus musculus. | | | |
| SOURCE | | Mus musculus | | | |
| ORGANISM | | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | | |
| REFERENCE | | 1 (bases 1 to 1438) | | | |
| AUTHORS | | Ju, T., Brewer, K., D'Souza, A., Cummings, R.D. and Canfield, W.M. | | | |
| TITLE | | Cloning and Expression of Human Core 1 beta | | | |
| JOURNAL | | J. Biol. Chem. 277 (1), 178-186 (2002) | | | |
| PUBMED | | 11677243 | | | |
| REFERENCE | | 2 (bases 1 to 1438) | | | |
| AUTHORS | | Ju, T.Z., Cummings, R.D. and Canfield, W.M. | | | |
| TITLE | | Cloning and characterization of core1 | | | |
| JOURNAL | | UDP-galactose:N-acetylgalactosamine-alpha-R beta | | | |
| REFERENCE | | 3 (bases 1 to 1438) | | | |
| AUTHORS | | Ju, T.Z., Cummings, R.D. and Canfield, W.M. | | | |
| TITLE | | Direct Submission | | | |
| JOURNAL | | | | | |
| FEATURES | | Location/Qualifiers | | | |
| source | | 1. .1438 | | | |
| | | /organism="Mus musculus" | | | |
| | | /db_xref="taxon:10090" | | | |
| Query Match. | | 59.8%; Score 292.6; DB 6; Length 1092; | | | |
| Best Local Similarity | | 76.5%; Pred. No. 1.1e-57; | | | |

| | | | |
|---------------------------|--|---|---|
| gene | 1. .1438 | /gene="C1galtl" | /function="responsible for the synthesis of core1 |
| CDS | 149. .1240 | /gene="C1galtl" | structure of O-glycans" |
| | /EC_number="2.4.1.122" | | /codon_start=1 |
| | /product="core1 | | UDP-galactose:N-acetylgalactosamine-alpha-R beta |
| | 1,3-galactosyltransferase" | | |
| | /protein_id="AAF81982.1" | | |
| | /db_xref="GI:8927164" | | |
| | /translation="MASKMLNLFVLCGSAIGFFLCQSLLSILLREEAAIQPMILHN | | |
| | DPHARHSDGHSHLKGKMFNFADSSQHKDENIDVAENLYQKVILCWYTSPOKLEK | | |
| | KAKAVKATWAGKRVKLFMSSEENQDPTVGLATKEGREOLYWKTKAQYVHDHYLE | | |
| | DADFMKADADTYIVIVDNLWLLSKYNPEQPIYVGRFRKPYVKGYMSGAGVLSKE | | |
| | ALRFEVNAFFTEKTKTHSSSIDIALGRCEMIIINVEAGDSRDTIGKTFPFPVPEHHLI | | |
| | KYQLPKPTFWLWNNYVPIRGPGCCSDIAVSFHYVDGTTMYELEYLVYHLRPPVGYLYR | | |
| | GYPALPENILKEINQVNRKEDTKIKUGNP" | | |
| misc_feature | 161. .241 | /gene="C1galtl" | |
| | /note="Region: non-cleavable transmembrane signal" | | |
| | /evidence=not_experimental | | |
| BASE COUNT | 465 a 286 c 298 g 389 t | | |
| ORIGIN | | | |
| Query Match | 59.8%; Score 292.6; DB 10; Length 1438; | | |
| Best Local Similarity | 76.5%; Pred. No. 1e-57; | | |
| Matches 374; Conservative | 0; Mismatches 109; Indels 6; Gaps 1; | | |
| QY | 1 | CATCTAAAAAGACTGATGAAGTTGATTGCAAAATGCTAGTCATCATATAATACGACACACA | 60 |
| Db | 320 | CACCTCAAGGACAGATCACTCAATGAGATTCAGCCACATAAAGATGAGAACAATA | 379 |
| QY | 61 | GGTGCTACTGACAAACCTCTATCAAAAAGATGAAAATTTCTTTGCTGGATTATGACAGGACCT | 120 |
| Db | 380 | GACGTTGCTGAGAACCTCTATCAGAAAGTTAAAATTTCTTTGCTGGATTATGACAGGACCT | 439 |
| QY | 121 | CAAAATCTGAAAAAAGATGACAGCATCAGAGATACATGGCCCGGAGGTTCGATATAA | 180 |
| Db | 440 | CARAATCTAGAAAMAGGCCCAACATGTCAAAAGCTACGTGGGCCCGACGGTGTATAA | 499 |
| QY | 181 | CGGTGTTGTTATGAGCTCAAAAAGAAAATAAGACTTCTCTACTGTGGGATTACACACCAA | 240 |
| Db | 500 | GTGTTATTTATGAGTTCGGAAGAAAATCAAGACTTCCCTACTGTGGGATTGAAACCAA | 559 |
| QY | 241 | GAAGACAGAACCAACTGCTCGAAAATAGTTAAAGCTTTTCTATATGCTCATGACCAT | 300 |
| Db | 560 | GAAGGACAGAGCAACTATTGGAAAACAATTAAGCTTTCCAGTATGTATCATGACCAT | 619 |
| QY | 301 | TATCTGGAATACATGATTGGTTGTTTCATGAAAGCAGATGATGATATATATATATACACA | 360 |
| Db | 620 | TATTTAGAAGATGCTGACTGGTTTATGAAAGCAGATGACGA-----CACATACGCTCAT | 673 |
| QY | 361 | TTGACAACCTTGAATGGCTTCTCACAACTATACCTGTATGAATCCACTTACTTTGGG | 420 |
| Db | 674 | GTGGACAACCTGAGATGGGTTCTTCAAAAGTATAACCTTGAAACAACCCATTTACTTTGGG | 733 |
| QY | 421 | AAAAGATTTAAGCACTGCGAGAAAACAGGACTACATGACTGGAGGAGCAGGATATGTACTG | 480 |
| Db | 734 | CGAAGATTTAAGCCCTATTGTGAAGCAGGGATACATAGCGGAGGACGGGCTATGTGCTCA | 793 |
| QY | 481 | AGCAAGAA 489 | |
| Db | 794 | AGCAAGAA 802 | |
| RESULT 14 | | | |
| AX175164 | | | |
| LOCUS | AX175164 | 1469 bp | DNA linear PAT 03-JUL-2001 |
| DEFINITION | Sequence 6 from Patent WO0144478. | | |

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 22:29:21 ; Search time 192.373 Seconds
(without alignments)
5724.434 Million cell updates/sec

Title: US-10-029-359A-3
Perfect score: 489
Sequence: 1 catctaaagactgatgaa.....gatagtactgagcaagaa 489

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| 1 | 489 | 100.0 | 489 | 24 | Human DNA encoding |
| 2 | 460.6 | 94.2 | 795 | 22 | Human breast cancer |
| 3 | 397.6 | 81.3 | 513 | 22 | Human breast cancer |
| 4 | 307 | 62.8 | 1794 | 22 | Human core 1 beta3 |
| 5 | 307 | 62.8 | 1800 | 21 | Nucleotide sequenc |
| 6 | 305.4 | 62.5 | 1560 | 24 | Human DNA encoding |
| 7 | 300.6 | 61.5 | 1440 | 22 | Rat core 1 beta3-g |
| 8 | 292.6 | 59.8 | 1469 | 22 | Mouse core 1 beta3 |
| 9 | 278.6 | 57.0 | 1424 | 23 | DNA encoding novel |

| | | | | | | |
|------|-------|------|--------|----|----------|--------------------|
| c 10 | 278 | 56.9 | 1975 | 22 | AAI36608 | Probe #5294 used t |
| c 11 | 278 | 56.9 | 1975 | 24 | ABS05361 | Human genome-deriv |
| c 12 | 277.4 | 56.7 | 668 | 22 | AAI49726 | Probe #18412 used |
| c 13 | 277.4 | 56.7 | 668 | 24 | ABS17860 | Human genome-deriv |
| c 14 | 168 | 34.4 | 505 | 24 | ABK88407 | Human DNA encoding |
| c 15 | 105 | 21.5 | 1172 | 22 | AAF90510 | Caenorhabditis ele |
| c 16 | 92.2 | 18.9 | 1167 | 22 | AAF90512 | Drosophila core 1 |
| c 17 | 92.2 | 18.9 | 1432 | 23 | ABL20061 | Drosophila melanog |
| c 18 | 92.2 | 18.9 | 1519 | 23 | ABL20059 | Drosophila melanog |
| c 19 | 92.2 | 18.9 | 1727 | 23 | ABL15335 | Drosophila melanog |
| c 20 | 92.2 | 18.9 | 4660 | 23 | ABL20060 | Drosophila melanog |
| c 21 | 92.2 | 18.9 | 9262 | 23 | ABL15334 | Drosophila melanog |
| c 22 | 92.2 | 18.9 | 9262 | 23 | ABL20058 | Drosophila melanog |
| c 23 | 90.6 | 18.5 | 2043 | 23 | ABL23893 | Drosophila melanog |
| c 24 | 79.6 | 16.3 | 1158 | 23 | ABL05467 | Drosophila melanog |
| c 25 | 74.8 | 15.3 | 1104 | 22 | AAF90511 | Drosophila core 1 |
| c 26 | 74.8 | 15.3 | 1474 | 23 | ABL05129 | Drosophila melanog |
| c 27 | 74.8 | 15.3 | 3641 | 23 | ABL05128 | Drosophila melanog |
| c 28 | 60.6 | 12.4 | 4553 | 23 | ABL23892 | Drosophila melanog |
| c 29 | 54.4 | 11.1 | 4058 | 23 | ABL12932 | Drosophila melanog |
| c 30 | 54 | 11.0 | 1310 | 23 | ABL12933 | Drosophila melanog |
| c 31 | 51.4 | 10.5 | 3442 | 23 | ABL05466 | Drosophila melanog |
| c 32 | 49.2 | 10.1 | 1672 | 23 | ABL20549 | Drosophila melanog |
| c 33 | 49.2 | 10.1 | 3442 | 23 | ABL05466 | Drosophila melanog |
| c 34 | 49.2 | 10.1 | 3814 | 23 | ABL20548 | Drosophila melanog |
| c 35 | 41.8 | 8.5 | 7244 | 22 | AA546726 | Tumour suppressor |
| c 36 | 41.6 | 8.5 | 1827 | 24 | ABN92451 | Staphylococcus epi |
| c 37 | 41.6 | 8.5 | 3003 | 22 | AAH54803 | S. epidermidis gen |
| c 38 | 38.8 | 7.9 | 60 | 24 | ABN42559 | Human spliced tran |
| c 39 | 38.6 | 7.9 | 640861 | 24 | ABA92787 | Buchnera sp. genom |
| c 40 | 37.6 | 7.7 | 3714 | 22 | AAH54411 | S. epidermidis gen |
| c 41 | 36.8 | 7.5 | 6095 | 22 | AA546309 | Tumour suppressor |
| c 42 | 36.8 | 7.5 | 6095 | 24 | ABL70149 | Chemically treated |
| c 43 | 36.8 | 7.5 | 6095 | 24 | ABL32360 | Human immune syste |
| c 44 | 36.8 | 7.5 | 6095 | 24 | ABL34474 | Human metastasis a |
| c 45 | 36.8 | 7.5 | 33053 | 24 | ABQ67006 | Human angiogenesis |

ALIGNMENTS

RESULT 1
ABK88408
ID ABK88408 standard; DNA; 489 BP.
AC
XX
AC
ABK88408;
XX
DT 07-OCT-2002 (first entry)
XX
XX
Human DNA encoding partial protein cluster II protein #2.
DE
XX
XX
Human; ds; gene; protein cluster II; obesity; diabetes mellitus;
KW
KW
XX
XX
Homo sapiens.
FH
FH
XX
XX
Key Location/Qualifiers
CDS 55...489
FT /*tag= a
FT /product= "Protein cluster II protein #2"
FT /note= "No start or stop codon shown"

XX WPI: 2002-57368/61.
DR P-PSDB; ABG30852.
XX
PT New protein cluster II nucleic acids and polypeptides, useful in
PT diagnosing metabolic diseases such as obesity and diabetes, and in
PT identifying agents for treating such diseases
XX
XX
PS Claim 1; Page 21; 28pp; English.
XX
XX The invention relates to a new isolated nucleic acid comprising:
CC (a) a fully defined sequence, encoding a protein cluster II protein,
CC appearing as ABK88407 and ABK88408; (b) a sequence capable of hybridising
CC under stringent hybridisation conditions to a nucleotide sequence
CC complementary to the polypeptide coding region of a nucleic acid in (a);
CC or (c) a sequence which is degenerate as a result of the genetic code to
CC a nucleotide sequence in (a) or (b). Also included are a isolated
CC an polypeptide encoded by the nucleic acid, a vector harbouring the
CC nucleic acid, a replicable expression vector, which carries and is
CC capable of mediating the expression of the nucleic acid a cultured host
CC cell harbouring the expression vector, producing a polypeptide by
CC culturing the cell, where the polypeptide is produced, and recovering the
CC polypeptide and identifying an agent capable of modulating the nucleic
CC acid by providing a cell comprising the nucleic acid molecule, contacting
CC the cell with a candidate agent, and monitoring the cell for an effect
CC that is not present in the absence of the candidate agent. The protein
CC cluster II nucleic acid and polypeptide is useful in the diagnosis of
CC metabolic diseases such as obesity and diabetes, and central nervous
CC system disorders and in the identification of agents for treating these
CC diseases. The nucleic acids may be used as hybridisation probes, for
CC chromosome and gene mapping, in polymerase chain reaction (PCR)
CC technologies, in the production of sense and antisense nucleic acids, and
CC in screening for new therapeutic molecules. The present sequence
CC encodes a protein cluster II protein of the invention.
XX
SQ Sequence 489 BP; 187 A; 86 C; 95 G; 121 T; 0 other;

Query Match 100.0%; Score 489; DB 24; Length 489;
Best Local Similarity 100.0%; Pred. No. 9e-124;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATCTAAAAGACTGATGAAGTTGATGCAAAATGCTAGTCATCATATAATACCAGAACACA 60
DB 1 CATCTAAAAGACTGATGAAGTTGATGCAAAATGCTAGTCATCATATAATACCAGAACACA 60

QY 61 GGTGTCACCTGACAACTCTATCAAAAGATGAAATTCCTTGTGCTATGATGACAGGACCT 120
DB 61 GGTGTCACCTGACAACTCTATCAAAAGATGAAATTCCTTGTGCTATGATGACAGGACCT 120

QY 121 CAAAATCTAGTAAAGATGACAGCATGATGATGATGATGATGATGATGATGATGATGATGAT 180
DB 121 CAAAATCTAGTAAAGATGACAGCATGATGATGATGATGATGATGATGATGATGATGATGAT 180

QY 181 GCGTTGTTTATGAGCTCAAAAGAAATAAAGACTTCTCTACTGTGGGATTACACACCAA 240
DB 181 GCGTTGTTTATGAGCTCAAAAGAAATAAAGACTTCTCTACTGTGGGATTACACACCAA 240

QY 241 GAAGACAGAAACCAACTGCTCGGAAATAGTAAAGCTTTCTATATGCTCATGACCAT 300
DB 241 GAAGACAGAAACCAACTGCTCGGAAATAGTAAAGCTTTCTATATGCTCATGACCAT 300

QY 301 TATCTGGAATACATGGATGGTTTCATGAAAGCAGATGATGATGATGATGATGATGATGATGAT 360
DB 301 TATCTGGAATACATGGATGGTTTCATGAAAGCAGATGATGATGATGATGATGATGATGATGAT 360

QY 361 TTGGACAACTTGAATGGCTTCTCAAACTATAACCCCTGATGATGATGATGATGATGATGATGAT 420
DB 361 TTGGACAACTTGAATGGCTTCTCAAACTATAACCCCTGATGATGATGATGATGATGATGATGAT 420

QY 421 AAAAGATTTAAGCACTGCGAGAAACAGGACTACATGCTGGAGGAGCAGGATATGATGATG 480
DB 421 AAAAGATTTAAGCACTGCGAGAAACAGGACTACATGCTGGAGGAGCAGGATATGATGATG 480

QY 481 AGCAAGAA 489
DB 481 AGCAAGAA 489

RESULT 2
AAL24100/c
ID AAL24100 standard; cDNA; 795 BP.
XX
XX
AC AAL24100;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 16557.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US00798.
XX
PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
XX WPI: 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer
XX
XX Claim 1; Page 3031-3032; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX
SQ Sequence 795 BP; 202 A; 171 C; 151 G; 271 T; 0 other;

Query Match 94.2%; Score 460.6; DB 22; Length 795;
Best Local Similarity 99.0%; Pred. No. 5.9e-116;
Matches 474; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1 CATCTAAAAGACTGATGAAGTTGATGCAAAATGCTAGTCATCATATAATACCAGAACACA 60
DB 481 CATCTAAAAGACTGATGAAGTTGATGCAAAATGCTAGTCATCATATAATACCAGAACACA 422

QY 61 GGTGTCACCTGACAACTCTATCAAAAGATGAAATTCCTTGTGCTATGATGACAGGACCT 120
DB 421 GGTGTCACCTGACAACTCTATCAAAAGATGAAATTCCTTGTGCTATGATGACAGGACCT 362

QY 121 CAAAATCTAG-AAAAAAGATGACAGCATGATGATGATGATGATGATGATGATGATGATGATG 179
DB 361 CAAAATCTAGAAAAAAGATGACAGCATGATGATGATGATGATGATGATGATGATGATGATG 302

QY 180 AGCGTTGTTTATGAGCTCAAAAGAAATAAAGACTTCTCTACTGTGGGATTACACACCAA 239
DB 180 AGCGTTGTTTATGAGCTCAAAAGAAATAAAGACTTCTCTACTGTGGGATTACACACCAA 239

Db 301 ACGGTTGTTATGAGCTCAAAAGAAAATAAAGACTTCTATATCTGTGGGATTACACACCAA 242
 QY 240 AGAAGACAGAAACCAACTGCTCTGGAATAAGTAAAGCTTTTCTATATGCTCATGACCA 299
 Db 241 AGAAGACAGAAACCAACTGCTCTGGAATAAGTAAAGCTTTTCTATATGCTCATGACCA 182
 QY 300 TTATCTGGATACATGAGTGGTTTCATGAAGCAGATGATGATATATATATATACAC 359
 Db 181 TTATCTGGATACATGAGTGGTTTCATGAAGCAGATGATGATATATATATATATACAC 122
 QY 360 ATTGACAACTTGAATGGCTTCTCACAACCTATACCTGATGAATCCACTTACTTTGG 419
 Db 121 ATTGACAACTTGAATGGCTTCTCACAACCTATACCTGATGAATCCACTTACTTTGG 62
 QY 420 GAAAGATTTAAGCACTGAGAAAACAGGACTACATGATGAGGAGGAGGATATGTAC 478
 Db 61 GAAAGATTTAAGCACTGAGAAAACAGGACTACATGATGAGGAGGAGGATATGTAC 3

RESULT 3

AAL15254/c

ID AAL15254 standard; cDNA; 513 BP.

XX AC AAL15254;

XX DT 07-DEC-2001 (first entry)

XX DE Human breast cancer expressed polynucleotide 7711.
 XX KW Human; breast cancer; cell marker; cytostatic; ss.
 XX OS Homo sapiens.

XX PN WO200151628-A2.

XX PD 19-JUL-2001.

XX PF 10-JAN-2001; 2001WO-US00798.

XX PR 14-JAN-2000; 2000US-0176077.

XX PR 14-MAR-2000; 2000US-0189167.

XX PR 24-MAR-2000; 2000US-0192099.

XX PR 29-MAR-2000; 2000US-0193480.

XX PR 15-MAY-2000; 2000US-0205230.

XX PR 09-JUN-2000; 2000US-0211315.

XX PR 25-JUL-2000; 2000US-0220534.

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Lillie J, Xu Y, Wang Y, Steinmann K;

XX DR WPI; 2001-451856/48.

XX PS New peptide useful as a marker for the diagnosis of breast cancer -
 XX Claim 1; Page 1386; 3695pp; English.

XX CC The invention relates to human breast cancer expressed polynucleotides
 CC (AAL07544-AAL26789) and methods of assessing whether a patient is
 CC afflicted with breast cancer by examining the correlation between the
 CC expression of certain markers and the cancerous state of breast cells.
 CC The polynucleotides and encoded polypeptides are potential markers for
 CC detecting, diagnosing, monitoring, characterising treating and
 CC potentially preventing breast cancer. The polynucleotides and encoded
 CC polypeptides are also useful for isolating compounds with cytostatic
 CC activity.

SQ Sequence 513 BP; 133 A; 90 C; 87 G; 180 T; 23 other;

Query Match 81.3%; Score 397.6; DB 22; Length 513;
 Best Local Similarity 93.8%; Pred. No. 8e-99;
 Matches 435; Conservative 0; Mismatches 24; Indels 5; Gaps 3;

QY 1 CATCTAAAAAGACTGATG-AAAGTTGTTGCAAAATGCTAGTCATCA--TAAATACCAGAAC 57
 Db 464 CATCTAAAAAGANINATNTAAGTTGTTGCAAAATGCTAGTCATCAATAAAATTCAGAAC 405
 QY 58 ACAGGTGCTACTGACAAAACCTCTATCAAAAGATGAAAATCTTTTGGTGGATTATGACAGGA 117
 Db 404 ACAGGTGCTACTGCAAAACCTCTATCAAAANATGAAAATCTTTTGGTGGATTATGNCAGSN 345
 QY 118 CCTCAAAATCT--AGAAAAAAGATCAGACCCATCAGAGATACATGGCCCCAGGGTTGCA 175
 Db 344 CCTCAAAATCTTAGAANAAAAAGATCAGACGNATCAGAGATACATGGGCCCCAGGGTTGCA 285
 QY 176 ATAAAGCGTTGTTATGAGCTCAAAAGAAAAATAAGACTTCTCTACTGTGGGATTACACA 235
 Db 284 ATAAAGCGTTGTTATGAGCTCAAAAGAAAAATAAGACTTCTCTACTGTGGGNTNACACA 225
 QY 236 CCAAGAAGACAGAAACCACTGCTCTGGAATAAGTAAAGCTTTTCTATATGCTCATG 295
 Db 224 CCAANAAGACAGAAACCACTGCTCTGGANATAGTTANAGCNTTCTATATGCTCTG 165
 QY 296 ACCATTATCTGGAATACATGATGGTTTCATGAAGCAGATGATGATATATATATATA 355
 Db 164 ACCATTATCTGGAATACATGATGGTTTCATGAAGCAGATGATGATATATATATA 105
 QY 356 TCACATTGGACAACTTGAATGGCTTCTCAAAAATAAACCTGATGAATCCACTTACT 415
 Db 104 TCACATTGGACAACTTGAATGGCTTCTCAAAAATAAACCTGATGAATCCACTTACT 45
 QY 416 TTGGAAAAAGATTTAAGCACTGCGAGAAACAGGACTACATGACT 459
 Db 44 TTGGAAAAAGATTTAAGCACTGCGAGAAACAGGACTACATGACT 1

RESULT 4

AAF90507

ID AAF90507 standard; cDNA; 1794 BP.

XX AC AAF90507;

XX DT 22-AUG-2001 (first entry)

XX DE Human core 1 beta3-galactosyl transferase cDNA.

XX KW Core 1 beta3-galactosyl transferase; human; O-glycosylation;
 KW galactosylation; glycosulfopeptide; Tn syndrome; IgA nephropathy;
 KW diagnosis; therapy; ss.

XX OS Homo sapiens.

XX FT Key Location/Qualifiers

XX CDS 63..1154

XX FT /*tag= a

XX FT /note= "a polynucleotide comprising this
 FT coding sequence is also claimed in Claim
 FT 1(A)"

XX WO200144478-A2.

XX PN 21-JUN-2001.

XX PD 14-DEC-2000; 2000WO-US33945.

XX PF 15-DEC-1999; 99US-0461321.

XX PR 15-DEC-1999; 99US-0464035.

XX PA (UYOK-) UNIV OKLAHOMA.

XX PI Canfield WM, Cummings RD, Ju T;

XX WPI; 2001-398157/42.

XX DR P-PSDB; AAB82456.

XX Novel purified core 1 beta3 galactosyl transferases of human, rat,
 PT

| | | |
|----|---|--|
| PI | Canfield WM, Cummings RD, Ju T; | |
| XX | WPI; 2001-398157/42. | |
| DR | P-PSDB; AAB82458. | |
| XX | | |
| PT | Novel purified core 1 beta3 galactosyl transferases of human, rat, | |
| PT | mouse, Drosophila melanogaster, Caenorhabditis elegans, useful for | |
| PT | galactosylating N-acetyl galactosamine linked to O-linking amino acid | |
| PT | on protein - | |
| XX | | |
| PS | Claim 1(A); Page 64; 74pp; English. | |
| XX | | |
| CC | The present sequence is that of mouse cDNA for core 1 | |
| CC | beta3-galactosyltransferase (b3Gtase, see AAB82458). The mouse | |
| CC | core 1 b3gtase has 89% identity and 94% similarity to the human | |
| CC | enzyme (see AAB82456). The invention provides human, rat, mouse, | |
| CC | Drosophila melanogaster and Caenorhabditis elegans b3Gtases and | |
| CC | nucleic acids encoding them. The polypeptides exhibit a wide | |
| CC | range of homologies. The polynucleotides can be used to transform | |
| CC | or transfect host cells for producing substantially pure forms of | |
| CC | the enzyme, or for use in an expression system for core 1 O-linked | |
| CC | glycosylation of proteins or peptides produced within the | |
| CC | expression system. The expressed enzymes can be used to | |
| CC | galactosylate, via a beta3 linkage, an N-acetylgalactosamine linked | |
| CC | to a serine, threonine or other O-linking amino acid on peptides or | |
| CC | proteins. Core 1 b3gtases are also useful in the synthesis of | |
| CC | glycosulfopeptides which can function as inhibitors of | |
| CC | P-selectin:PSG1-1 interactions. Other potential uses include | |
| CC | diagnostic tests for the rare Tn-syndrome or IgA nephropathy, and | |
| CC | therapy of these disorders. | |
| XX | | |
| SQ | Sequence 1469 BP; 467 A; 296 C; 313 G; 393 T; 0 other; | |

| | | | | | | | |
|-----------------------|--------|--------------|----------------------------|---------------------------------|------|--------|---------|
| Query Match | 59.8%; | Score | 292.6; | DB | 22; | Length | 1469; |
| Rest Local Similarity | 76.5%; | Pred. | No. 5.1e-70; | | | | |
| Matches | 374; | Conservative | 0; | Mismatches | 109; | Indels | 6; Gaps |
| 1 | 1 | CATCTAA | AAAAGACGTGATGAAGTTGATTCG | AAATGCTAGTCATCAATAATACCAGAACACA | 60 | | |
| | | | | | | | |
| 351 | 351 | CACCTCA | AGGACAGATGAATCTCAATG | CGAGATTCGAGCCAACTAAAGATGAGAACA | 410 | | |
| | | | | | | | |
| 61 | 61 | GGTGTCA | CTGCAGAACTCTATCAAAAGATCAAA | AAATCTTTGCTGGGATTAACACAGAGCT | 120 | | |
| | | | | | | | |
| 411 | 411 | GAGTGTG | CTGAGACCTCTATCAGAAAGTTAA | AAATTTCTTTGTTGGGTTAATGACAAGTCT | 470 | | |
| | | | | | | | |
| 121 | 121 | CAAAATC | TAGAAAAAAGATCAGACGATCAGAG | ATACATGGGCCAGGGTTGCAATATAA | 180 | | |
| | | | | | | | |
| 471 | 471 | CAAAATC | TAGAGAAAAGGCCAACATGTCAA | AGCTACGTGGGCCAGCGTTGTATAA | 530 | | |
| | | | | | | | |
| 181 | 181 | CGGTTGT | TATGACCTCAAAAGAAATAAAGACT | CTCTACTGTGGGATTACACACCAAA | 240 | | |
| | | | | | | | |
| 531 | 531 | GGTGTAT | TATTTATGNGTTGCGAAGAAAATCA | AGACTTCCCTACTCTGGGATTGAAAAC | 590 | | |
| | | | | | | | |
| 241 | 241 | GAAGAC | AGAAACCAACTGTCTCGAAAAATAG | TAAAGCTTTTCTATATGCTCATGACC | 300 | | |
| | | | | | | | |
| 591 | 591 | GAAGGC | ACAGAGCAACTATATTGGA | AAACAAATTAAGCTTCCAGTATGACATG | 650 | | |
| | | | | | | | |
| 301 | 301 | TATCTG | GAATACATGGATTGGTTTCATGAA | ACGAGATGATATATGATATATATACA | 360 | | |
| | | | | | | | |
| 651 | 651 | TATTTA | GAAAGATGCTGACTGGTTTATCAA | AGCAGATGACGA-----CACATACCT | 704 | | |
| | | | | | | | |
| 361 | 361 | TGGGAC | ACTTGAATGGCTTCTCAACA | ACTATAACCCCTGATGAATCCACTTTGG | 420 | | |
| | | | | | | | |
| 705 | 705 | GTGGAC | ACCTTGAGATGGCTTCTATCAA | AGTATAGCCCTGAAACACCCATTTACT | 764 | | |
| | | | | | | | |
| 421 | 421 | AAAAGAT | TTAAGCACTCCAGAAACAGGACT | ACATGACTGGAGGAGCAGGATATGCT | 480 | | |
| | | | | | | | |
| 765 | 765 | CGAAGT | TTAAGCCCTATGTGA | ACAGAGGATACATGCGCGGAGGCGGCT | 824 | | |
| | | | | | | | |
| 481 | 481 | AGCA | AGAA | 489 | | | |
| | | | | | | | |
| 825 | 825 | AGCA | AGAA | 833 | | | |
| | | | | | | | |
| db | | | | | | | |

| | | | |
|-----|----|--|-----|
| 248 | Db | GACATTGCTGAAAGCGCTCGAGGAGAAAGTTAGAAATCTTTTGCTGGGTTATGACCGGCC | 307 |
| 120 | Qy | TC-AAAACTAGAAAAAAGATCAGACGCATCAGAGATACATGGGCCAGGGTTGCAATA | 178 |
| 308 | Db | TCAAAACCTAGAGAAAAAGGCCAAACACGCTCAAGCTACTTTGGCCCCAGCGTTGTAACA | 367 |
| 179 | Qy | AGCGTTGTTTATGAGCTCAAAAGAAAATAAAGACTTCTCTACTGTGGGATTCACACACCA | 238 |
| 368 | Db | AAAGTGTGTTTATGAGTTTCAGAAGAAAATAAAGACTTCCCTGCTGTGGGACTGAAAAACCA | 427 |
| 239 | Qy | AAAGAACAAGAAACCACACTGCTCTGGAAAAATAGTTAAAGCTTTTCTATATGCTCATGACC | 298 |
| 428 | Db | AGAAGGCAGAGATCAACTATCTAGGAACAAATAAAGCTTTTCACTATGTCATGAAC | 487 |
| 299 | Qy | ATTATCTGGAATACATGAGTGGTTTCATGAAGCAGATGATATATATATATATATCA | 358 |
| 488 | Db | ATTATTTATAAGATGCTGATTGGTTTTTGAAGCAGATGATGA-----CACGTATGTCA | 541 |
| 359 | Qy | CATTGGACAACTTGAATGGCTTCTCAAAACTATATACCCTGATGAACTCACTTACTTTG | 418 |
| 542 | Db | TACTAGACAATTTGAGGTGGCTTCTTCAAATACGACCTCGAAGAACCCATTACTTTG | 601 |
| 419 | Qy | GGAAAGATTTTAAGCACCTCGCAAAAACAGGACTACATGACTTGGAGGACGAGGATATGTAC | 478 |
| 602 | Db | GGAGAAGATTTAAGCCTTATGTAAGCAGCGCTACATGAGTGGAGGACGAGGATATGTAC | 661 |
| 479 | Qy | TGAGCAAGAA 489 | |
| 662 | Db | TAAGCAAGAA 672 | |

| | |
|------------|---|
| RESULT 10 | |
| AAI36608/c | |
| ID | AAI36608 standard; DNA; 1975 BP. |
| XX- | |
| XX | AAI36608; |
| XX | |
| XX | 17-OCT-2001 (first entry) |
| DE- | Probe #3294 used to measure gene expression in human placenta sample. |
| XX | |
| KW | Probe; microarray; human; placenta; antenatal diagnosis; |
| KW | genetic disorder; ss. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| PN | WO200157272-A2. |
| XX | |
| PD | 09-AUG-2001. |
| XX | |
| PF | 30-JAN-2001; 2001WO-US00663. |
| XX | |
| XX | 04-FEB-2000; 2000US-0180312. |
| PR | 26-MAY-2000; 2000US-0207456. |
| PR | 30-JUN-2000; 2000US-0608408. |
| PR | 03-AUG-2000; 2000US-0632366. |
| PR | 21-SEP-2000; 2000US-0234687. |
| PR | 27-SEP-2000; 2000US-0236359. |
| PR | 04-OCT-2000; 2000GB-0024263. |
| XX | |
| PA | (MOLE-) MOLECULAR DYNAMICS INC. |
| XX | |
| PI | Penn SG, Hanzel DK, Chen W, Rank DR; |
| XX | |
| DR | WPI; 2001-488897/53. |
| XX | |
| PT | Human genome-derived single exon nucleic acid probes useful for |
| XX | analyzing gene expression in human placenta - |
| XX | |
| PS | Claim 25; SEQ ID No 5294; 654pp; English. |
| XX | |
| CC | The present invention relates to single exon nucleic acid probes (SENPs). |
| CC | The present sequence is one such probe. The probes are useful for |

| | | | | | | |
|------------|--|--|--------------------|----------------|--------------|---------|
| CC | producing a microarray for predicting, measuring and displaying gene | | | | | |
| CC | expression in samples derived from human placenta. The probes are useful | | | | | |
| CC | for antenatal diagnosis of human genetic disorders. | | | | | |
| XX | | | | | | |
| SQ | Sequence 1975 BP; 627 A; 395 C; 347 G; 606 T; 0 other; | | | | | |
| | Query Match | 56.9%; | Score 278; | DB 22; | Length 1975; | |
| | Best Local Similarity | 78.4%; | Pred. No. 5.4e-66; | | | |
| | Matches 348; | Conservative | 0; | Mismatches 90; | Indels 6; | Gaps 1; |
| QY | 46 | AAATACCAAGAACACAGGTGTCCTACTGCACAACACTCTATCAAAGAAGATGAATAATTCTTTTGGCTGG | 105 | | | |
| DB | 938 | ATAGATCAGAAACACAGACATGCTGTAACACCTCTATCAGAAAGTTAGAAATCTTTTGCTGG | 879 | | | |
| QY | 106 | ATTATGACAGGACCTCAAAATCTAGAAAAAAGATCAGACGCATCAGAGATACATGGGCC | 165 | | | |
| DB | 878 | GTTATGACCGCGCCTCAAAAACCTAGAGAAAAAGGCCAACACGCTCAAAGCTACTTTGGGCC | 819 | | | |
| QY | 166 | CAGGTTTGCAATAAAGCGTTGTTTATCAGCTCAAAAGAAAAATAAAGACTTCTCTACTGTG | 222 | | | |
| DB | 818 | CACGTTGTAAACAAAGTGTGTTATGAGTTCAGAGAAAAATAAAGACTTCCCTGCTGTG | 759 | | | |
| QY | 226 | GGATTACACACCAGAAAGACAGAAACCAACTCTCTGGAATAAGTAAAGCTTTTTCTA | 285 | | | |
| DB | 758 | GGACTGAAACACCAAGAACGACAGAGATCAACTATACTGGAACCAATATAAGCTTTTTTCAG | 699 | | | |
| QY | 286 | TATGCTCATCACCAATATCTGGAATACATCGATTGGTTTCATGAAGCAGATGATGATATA | 345 | | | |
| DB | 698 | TATGTTTCATGAACATATTTAGAAGAAGTCGATTGGTTTTTGAAGCAGATGATGA---- | 643 | | | |
| QY | 346 | TGTATATATATACATTTGGACAACTTCGAAATGGCTTCTCAAACTATAACCTGATGAA | 405 | | | |
| DB | 642 | --CACGTATCTACTACTAGACAATTTGAGCTGGCTCTTTTCAAAATACGACCTGAAGAA | 585 | | | |
| QY | 406 | TCACATTACTTTGGGAAAAAGATTTAAGCACTGCAGAAACAGGACTACATGCTGGAGGA | 465 | | | |
| DB | 584 | CCCATTACTTTGGGAGAAGATTTAAGCCTTATTAAGCAGGGCTACATGAGTGGAGGA | 525 | | | |
| QY | 466 | GCAGGATATGCTACTGAGCAAGAA | 489 | | | |
| DB | 524 | GCAGGATATGCTACTAAGCAAGAA | 501 | | | |
| RESULT 11 | | | | | | |
| ABS05361/c | | | | | | |
| ID | ABS05361 standard; DNA; 1975 BP. | | | | | |
| XX | ABS05361; | | | | | |
| AC | ABS05361; | | | | | |
| XX | 19-AUG-2002 (first entry) | | | | | |
| DT | Human genome-derived single exon probe from lung SEQ ID NO 5352. | | | | | |
| DE | Human; ds; single exon probe; asthma; lung cancer; COPD; ILD; | | | | | |
| KW | chronic obstructive pulmonary disease; interstitial lung disease; | | | | | |
| KW | familial idiopathic pulmonary fibrosis; neurofibromatosis; | | | | | |
| KW | tuberosus sclerosis; Gaucher's disease; Niemann-Pick disease; | | | | | |
| KW | Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis; | | | | | |
| KW | pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome; | | | | | |
| KW | pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia; | | | | | |
| KW | primary ciliary dyskinesia; pulmonary hypertension; | | | | | |
| KW | hyaline membrane disease. | | | | | |
| OS | Homo sapiens. | | | | | |
| XX | WO200186003-A2. | | | | | |
| PN | 15-NOV-2001. | | | | | |
| PD | 30-JAN-2001; 2001WO-US00665. | | | | | |
| XX | 04-FEB-2000; 2000US-180312P. | | | | | |
| PR | 26-MAY-2000; 2000US-207456P. | | | | | |

PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-234687P.
 PR 27-SEP-2000; 2000US-236359P.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2002-114183/15.
 XX Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples -
 XX Claim 1; SEQ ID No 5352; 634pp; English.
 PS
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of
 CC probes; the novel set of probes which hybridise at high stringency to a
 CC nucleic acid expressed in the human lung; measuring gene expression in a
 CC sample derived from human lung, comprising (a) contacting the array with
 CC a collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
 CC and hyaline membrane disease. The present sequence is a single exon
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1975 BP; 627 A; 395 C; 347 G; 606 T; 0 other;

Query Match 56.9%; Score 278; DB 24; Length 1975;
 Best Local Similarity 78.4%; Pred. No. 5.4e-66;
 Matches 348; Conservative 0; Mismatches 90; Indels 6; Gaps 1;
 46 AAATACAGAACACAGTGTCACTGACAACTCTATCAAAAGATGAAATCTTCTCTGG 105
 938 ATAGATGAGACACACACATGCTGAAACCTCTATCAGAAATGAAATCTTCTCTGG 879
 106 ATTATGACAGGACCTCAAAATCTTAAAGAAAGATCAGACGCGATCAGATCATCGGCC 165
 878 GTTATGACCGGCCCTCAAAACCTTAGAGAAAAGCCCAACACGCTCAAGCTACTTGGCC 819
 166 CAGGGTTGCATTAAGCGTTGTTTATGAGCTCAAAAGAAATAAGACTTCTCTACTGTG 225

Db 818 CAGGTTGTACAAAGTGTGTTTATGAGTTCAGAAAGAAATAAGACTTCCCTGCTGTG 759
 Qy 226 GGATTACACACCAGAAAGACAGAAACCAACTGTCTCTGGAAATAGTTAAAGCTTTTCTA 285
 Db 758 GGACTGAAACCAAGGAGAGATCAACTACTGTAAGAAACAATTAAGCTTTTCTAG 699
 Qy 286 TATGCTCATGACCATTTATCTGGATACATGGATGTTGTTTCAAGCAGATGATGATATA 345
 Db 698 TATGTTTCATGAACATATTTAGAGAGATGCTGATTTGTTTGAAGCAGATGATGA---- 643
 Qy 346 TGATATATATATCAATTTGGACAACTTGAATGGCTTCTCACAACATATAACCTGTATGAA 405
 Db 642 --CACGTATGTCATCTAGACATTTGAGGTGGCTTCTTCAAAATACGACCTGGAAGA 585
 Qy 406 TCACATTTACTTTGGGAAAAGATTTAAGCACTGTCAGAAAACAGGACTACATGACTGGAGGA 465
 Db 584 CCCATTTACTTTGGGAGAAGATTTAAGCCCTATGTAAGCAGGCGCTACATGAGTGGAGGA 525
 Qy 466 GCAGGATATGCTACTGACCAAGAA 489
 Db 524 GCAGGATATGCTACTAAGCAAGAA 501

RESULT 12
 AAI49726/C
 ID AAI49726 standard; DNA; 668 BP.
 XX
 AC AAI49726;
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Probe #18412 used to measure gene expression in human placenta sample.
 XX
 KW Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder; ss.
 XX Homo sapiens.
 OS
 XX
 PN W0200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US006663.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488897/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX
 PS Claim 25; SEQ ID No 18412; 654pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 XX
 SQ Sequence 668 BP; 178 A; 144 C; 115 G; 231 T; 0 other;
 Query Match 56.7%; Score 277.4; DB 22; Length 668;

| | | | | | | | | | | | | | | | | | |
|--|--|--|-----|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| Best Local Similarity 78.9%; Pred. No. 5.6e-66; Matches 345; Conservative 0; Mismatches 86; Indels 6; Gaps 1; | | | | | | | | | | | | | | | | | |
| QY | 53 | AGAACACAGGTGTCACCTGACAAACTCTATCAAAAGATGAAAAATCTTTTCTGCTGGATTATGA | 112 | | | | | | | | | | | | | | |
| Db | 665 | AGAACACAGACATGCTGTAACCTCTATCAGAAAGTTAGAAATCTTTTCTGCTGGATTATGA | 606 | | | | | | | | | | | | | | |
| QY | 113 | CAGGACCTCAAAATCTAGAAAAAAGATCAGAGCATCAGAGATACATGCGCCACAGGGTT | 172 | | | | | | | | | | | | | | |
| Db | 605 | CCGGCCCTCAAAACCTAGAGAAAAAGGCCAACACAGCTCAAAAGCTACTTGGGCCACGGTT | 546 | | | | | | | | | | | | | | |
| QY | 173 | GCAATAAAGCGTTGTTATGAGCTCAAAAGAAATAAAGACTTCTTCTACTGTGGATTATC | 232 | | | | | | | | | | | | | | |
| Db | 545 | GTAACAAAGTGTGTTTATGAGTTCAAGAAAAATAAAGACTTCCCTGCTGTGGGACTGA | 486 | | | | | | | | | | | | | | |
| QY | 233 | ACACCAAGAAGACAGAAACCAACTGCTCTGGAAATAGTTAAAGCTTTTCTATATGCTC | 292 | | | | | | | | | | | | | | |
| Db | 485 | AAACCAAGAAGGAGAGATCACTATATCTGGAACAAATAAAGCTTTTCACTATATGCTC | 426 | | | | | | | | | | | | | | |
| QY | 293 | ATGACCATTTCTGGAATACATGATTTGTTTCATGAAAGCAGATGATATATGATAT | 352 | | | | | | | | | | | | | | |
| Db | 425 | ATGAACATTTATGAAAGATGCTGATTTGTTTGAAGCAGATGATGA-----CACGT | 372 | | | | | | | | | | | | | | |
| QY | 353 | ATATCACATTTGGCAACTTGAATGGCTTCTCAAAACTATACCTGATGATCCACTT | 412 | | | | | | | | | | | | | | |
| Db | 371 | ATGTCATACTAGACAATTTGAGTGGCTTCTTTCAAAATACGACCTTGAAAGACCCATTT | 312 | | | | | | | | | | | | | | |
| QY | 413 | ACTTTGGGAAAGATTTAAGCACTGCAACAAAGCACTACATGCTGAGGAGCAGGAT | 472 | | | | | | | | | | | | | | |
| Db | 311 | ACTTTGGGAGAGATTTAAACCTTATGTAAGCAGGCTACATGAGTGGAGGAGCAGGAT | 252 | | | | | | | | | | | | | | |
| QY | 473 | ATGCTACTGACCAAGAA 489 | | | | | | | | | | | | | | | |
| Db | 251 | ATGCTACTAAGCAAGAA 235 | | | | | | | | | | | | | | | |
| RESULT 13 | | | | | | | | | | | | | | | | | |
| ID | ABS17860/c | | | | | | | | | | | | | | | | |
| XX | ABS17860 standard; DNA; 668 BP. | | | | | | | | | | | | | | | | |
| AC | ABS17860; | | | | | | | | | | | | | | | | |
| XX | 19-AUG-2002 (first entry) | | | | | | | | | | | | | | | | |
| DT | Human genome-derived single exon probe ORF from lung SEQ ID No 17851. | | | | | | | | | | | | | | | | |
| DE | Human; ds; single exon probe; asthma; lung cancer; COPD; ILD; | | | | | | | | | | | | | | | | |
| KW | chronic obstructive pulmonary disease; interstitial lung disease; | | | | | | | | | | | | | | | | |
| KW | familial idiopathic pulmonary fibrosis; neurofibromatosis; | | | | | | | | | | | | | | | | |
| KW | tuberosus sclerosis; Gaucher's disease; Niemann-Pick disease; | | | | | | | | | | | | | | | | |
| KW | Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis; | | | | | | | | | | | | | | | | |
| KW | pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome; | | | | | | | | | | | | | | | | |
| KW | pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia; | | | | | | | | | | | | | | | | |
| KW | primary ciliary dyskinesia; pulmonary hypertension; | | | | | | | | | | | | | | | | |
| KW | hyaline membrane disease; open reading frame; ORF. | | | | | | | | | | | | | | | | |
| OS | Homo sapiens. | | | | | | | | | | | | | | | | |
| XX | WO200186003-A2. | | | | | | | | | | | | | | | | |
| PN | 15-NOV-2001. | | | | | | | | | | | | | | | | |
| PD | | | | | | | | | | | | | | | | | |
| XX | 30-JAN-2001; 2001WO-US000665. | | | | | | | | | | | | | | | | |
| PR | 04-FEB-2000; 2000US-180312P. | | | | | | | | | | | | | | | | |
| PR | 26-MAY-2000; 2000US-207456P. | | | | | | | | | | | | | | | | |
| PR | 30-JUN-2000; 2000US-0608408. | | | | | | | | | | | | | | | | |
| PR | 03-AUG-2000; 2000US-0632366. | | | | | | | | | | | | | | | | |
| PR | 21-SEP-2000; 2000US-234687P. | | | | | | | | | | | | | | | | |
| PR | 27-SEP-2000; 2000US-236359P. | | | | | | | | | | | | | | | | |
| PR | 04-OCT-2000; 2000GB-0024263. | | | | | | | | | | | | | | | | |
| XX | (MOLE-) MOLECULAR DYNAMICS INC. | | | | | | | | | | | | | | | | |
| PA | | | | | | | | | | | | | | | | | |
| XX | Penn SG, Hanzel DK, Chen W, Rank DR; | | | | | | | | | | | | | | | | |
| PI | WPI; 2002-114183/15. | | | | | | | | | | | | | | | | |
| DR | | | | | | | | | | | | | | | | | |
| XX | | | | | | | | | | | | | | | | | |
| PT | Spatially-addressable set of single exon nucleic acid probes, used to | | | | | | | | | | | | | | | | |
| PT | measure gene expression in human lung samples - | | | | | | | | | | | | | | | | |
| XX | Claim 4; SEQ ID No 17851; 634pp; English. | | | | | | | | | | | | | | | | |
| PS | | | | | | | | | | | | | | | | | |
| XX | | | | | | | | | | | | | | | | | |
| CC | The invention relates to a spatially-addressable set of single exon | | | | | | | | | | | | | | | | |
| CC | nucleic acid probes for measuring gene expression in a sample derived | | | | | | | | | | | | | | | | |
| CC | from human lung comprising single exon nucleic acid probes having one of | | | | | | | | | | | | | | | | |
| CC | 12614 nucleic acid sequences mentioned in the specification, or their | | | | | | | | | | | | | | | | |
| CC | complements or the 12387 open reading frames derived from the 12614 | | | | | | | | | | | | | | | | |
| CC | probes. Also included are a microarray comprising the novel set of | | | | | | | | | | | | | | | | |
| CC | probes; the novel set of probes which hybridise at high stringency to a | | | | | | | | | | | | | | | | |
| CC | nucleic acid expressed in the human lung; measuring gene expression in a | | | | | | | | | | | | | | | | |
| CC | sample derived from human lung, comprising (a) contacting the array with | | | | | | | | | | | | | | | | |
| CC | a collection of detectably labeled nucleic acids derived from human lung | | | | | | | | | | | | | | | | |
| CC | mRNA, and (b) measuring the label detectably bound to each probe of | | | | | | | | | | | | | | | | |
| CC | the array; identifying exons in a eukaryotic genome, comprising | | | | | | | | | | | | | | | | |
| CC | (a) algorithmically predicting at least one exon from genomic sequences | | | | | | | | | | | | | | | | |
| CC | of the eukaryote; and (b) detecting specific hybridisation of detectably | | | | | | | | | | | | | | | | |
| CC | labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, | | | | | | | | | | | | | | | | |
| CC | having a fragment identical to the predicted exon, the probe is included | | | | | | | | | | | | | | | | |
| CC | in the above mentioned microarray; assigning exons to a single gene, | | | | | | | | | | | | | | | | |
| CC | comprising (a) identifying exons from genomic sequence by the method | | | | | | | | | | | | | | | | |
| CC | above and (b) measuring the expression of each of the exons in several | | | | | | | | | | | | | | | | |
| CC | tissues and/or cell types using hybridisation to a single exon | | | | | | | | | | | | | | | | |
| CC | microarrays having a probe with the exon, where a common pattern of | | | | | | | | | | | | | | | | |
| CC | expression of the exons in the tissues and/or cell types indicates that | | | | | | | | | | | | | | | | |
| CC | the exons should be assigned to a single gene; a peptide comprising one | | | | | | | | | | | | | | | | |
| CC | of 12011 sequences, mentioned in the specification, or encoded by the | | | | | | | | | | | | | | | | |
| CC | probes/open reading frames (ORF). The probes are used for gene | | | | | | | | | | | | | | | | |
| CC | expression analysis, and for identifying exons in a gene, particularly | | | | | | | | | | | | | | | | |
| CC | using human lung derived mRNA and for the study of lung diseases | | | | | | | | | | | | | | | | |
| CC | such as asthma, lung cancer, chronic obstructive pulmonary disease | | | | | | | | | | | | | | | | |
| CC | (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary | | | | | | | | | | | | | | | | |
| CC | fibrosis, neurofibromatosis, tuberosus sclerosis, Gaucher's disease, | | | | | | | | | | | | | | | | |
| CC | Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary | | | | | | | | | | | | | | | | |
| CC | haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis, | | | | | | | | | | | | | | | | |
| CC | pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic | | | | | | | | | | | | | | | | |
| CC | pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension | | | | | | | | | | | | | | | | |
| CC | and hyaline membrane disease. The present sequence is a single exon | | | | | | | | | | | | | | | | |
| CC | probe open reading frame of the invention. | | | | | | | | | | | | | | | | |
| CC | Note: The sequence data for this patent did not form part | | | | | | | | | | | | | | | | |
| CC | of the printed specification, but was obtained in electronic | | | | | | | | | | | | | | | | |
| CC | format directly from WIPO at | | | | | | | | | | | | | | | | |
| CC | ftp.wipo.int/pub/published_pct_sequences. | | | | | | | | | | | | | | | | |
| XX | | | | | | | | | | | | | | | | | |
| SQ | Sequence 668 BP; 178 A; 144 C; 115 G; 231 T; 0 other; | | | | | | | | | | | | | | | | |
| Query Match 56.7%; Score 277.4; DB 24; Length 668; | | | | | | | | | | | | | | | | | |
| Best Local Similarity 78.9%; Pred. No. 5.6e-66; | | | | | | | | | | | | | | | | | |
| Matches 345; Conservative 0; Mismatches 86; Indels 6; Gaps 1; | | | | | | | | | | | | | | | | | |
| QY | 53 | AGAACACAGGTGTCACCTGACAAACTCTATCAAAAGATGAAAAATCTTTTCTGCTGGATTATGA | 112 | | | | | | | | | | | | | | |
| Db | 665 | AGAACACAGACATGCTGTAACCTCTATCAGAAAGTTAGAAATCTTTTCTGCTGGATTATGA | 606 | | | | | | | | | | | | | | |
| QY | 113 | CAGGACCTCAAAATCTAGAAAAAAGATCAGAGCATCAGAGATACATGCGCCACAGGGTT | 172 | | | | | | | | | | | | | | |
| Db | 605 | CCGGCCCTCAAAACCTAGAGAAAAAGGCCAACACAGCTCAAAAGCTACTTGGGCCACGGTT | 546 | | | | | | | | | | | | | | |
| QY | 173 | GCAATAAAGCGTTGTTATGAGCTCAAAAGAAATAAAGACTTCTTCTACTGTGGATTATC | 232 | | | | | | | | | | | | | | |
| Db | 545 | GTAACAAAGTGTGTTTATGAGTTCAAGAAAAATAAAGACTTCCCTGCTGTGGGACTGA | 486 | | | | | | | | | | | | | | |
| QY | 233 | ACACCAAGAAGACAGAAACCAACTGCTCTGGAAATAGTTAAAGCTTTTCTATATGCTC | 292 | | | | | | | | | | | | | | |
| Db | 485 | AAACCAAGAAGGAGAGATCACTATATCTGGAACAAATAAAGCTTTTCACTATATGCTC | 426 | | | | | | | | | | | | | | |

QY 293 ATGACCAATTCTCGAATACATGATGGTTTCATGAAAGCAGATGATGATATATATAT 352
 Db IIII IIII I III IIII IIII IIII IIII IIII IIII IIII IIII IIII
 425 ATGAACATATTATGAAGATGCTGATGGTGTGTTTGAAGCAGATGATGA-----CACGT 372
 QY 353 ATATCACATTTGGACAACCTTGAATGGCTTCTCACAACATATACCCCTGATGATCCACTT 412
 Db IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
 371 ATGTCATCTAGACAAATTTGAGGTGGCTTCTTCAAAATACGACCCCTGAAGAACCCTATT 312
 QY 413 ACTTTGGGAAAGATTTAAGCACTGCGACAAACAGGACTACATGACTGGAGGAGCAGGAT 472
 Db IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
 311 ACTTTGGGAGAGATTTAAGCCCTTATGTAAGCAGGCTACATGATGGAGGAGCAGGAT 252
 QY 473 ATGTACTGAGCAAGAA 489
 Db IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
 251 ATGTACTAAGCAAGAA 235

RESULT 14

ID ABK88407 standard; DNA; 505 BP.
 XX AC

ABK88407;

07-OCT-2002 (first entry)

Human DNA encoding partial protein cluster II protein #1.

Human; ds; gene; protein cluster II; obesity; diabetes mellitus;
 central nervous system disorder; metabolic disease.

OS Homo sapiens.

XX Key Location/Qualifiers
 XX CDS 21..497

FT /*tag= a
 FT /product= "Protein cluster II protein #1"
 FT /note= "No stop codon shown"
 FT /transl_except= (pos:297,299,aa:Xaa)
 FT /note= "Xaa= unknown"

PN WO200251864-A1.

XX 04-JUL-2002.

XX 14-DEC-2001; 2001WO-SE02786.

XX 22-DEC-2000; 2000SE-0004828.

XX (PHAA) PHARMACIA AB.

XX Attersand A;

XX WPI; 2002-575368/61.

DR P-PSDB; ABG30851.

XX New protein cluster II nucleic acids and polypeptides, useful in
 PT diagnosing metabolic diseases such as obesity and diabetes, and in
 PT identifying agents for treating such diseases

PS Claim 1; Page 20-21; 28pp; English.

XX The invention relates to a new isolated nucleic acid comprising:
 CC (a) a fully defined sequence, encoding a protein cluster II protein,
 CC appearing as ABK88407 and ABK88408; (b) a sequence capable of hybridising
 CC under stringent hybridisation conditions to a nucleotide sequence
 CC complementary to the polypeptide coding region of a nucleic acid in (a);
 CC or (c) a sequence which is degenerate as a result of the genetic code to
 CC a nucleotide sequence in (a) or (b). Also included are a isolated
 CC an polypeptide encoded by the nucleic acid, a vector harbouring the
 CC nucleic acid, a replicable expression vector, which carries and is
 CC capable of mediating the expression of the nucleic acid a cultured host
 CC cell harbouring the expression vector, producing a polypeptide by
 CC culturing the cell, where the polypeptide is produced, and recovering the

CC polypeptide and identifying an agent capable of modulating the nucleic
 CC acid by providing a cell comprising the nucleic acid molecule, contacting
 CC the cell with a candidate agent, and monitoring the cell for an effect
 CC that is not present in the absence of the candidate agent. The protein
 CC cluster II nucleic acid and polypeptide is useful in the diagnosis of
 CC metabolic diseases such as obesity and diabetes, and central nervous
 CC system disorders and in the identification of agents for treating these
 CC diseases. The nucleic acids may be used as hybridisation probes, for
 CC chromosome and gene mapping, in polymerase chain reaction (PCR)
 CC technologies, in the production of sense and antisense nucleic acids, and
 CC in screening for new therapeutic molecules. The present sequence
 CC encodes a protein cluster II protein of the invention.

XX Sequence 505 BP; 166 A; 95 C; 108 G; 135 T; 1 other;

Query Match 34.4%; Score 168; DB 24; Length 505;

Best Local Similarity 74.7%; Pred. No. 3.7e-36; Indels 0; Gaps 0;
 Matches 210; Conservative 0; Mismatches 71;

QY 1 CATCTAAAAAGACTGATGAAGTTGATTGCAAAATGCTAGTCATCATATAATACCAACACACA 60
 Db IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII

216 CATCTAGGAGGACAAATGAACTTCAATGCAGATCTTAGCCCAACCTAAGATGAGACACA 275
 Db IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII

QY 61 GGTGTCACCTCACAACTCTATCAAAAGATGAAATTTCTTGTGGATATGACAGGACCT 120
 Db IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII

276 GAAATCGCTGAAACCTCTATNAGCAAGTTAAAAATTTCTTGTGGTTATGACAGGCTCT 335
 Db IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII

QY 121 CAAATCTAGAAAAAGATCAGACGCTCAGAGATACATGGCCCCAGGTTGCAATAAA 180
 Db IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII

336 CAAAACCTCAGAAAAAGGCCAACATGTCAAAGCTACATGGGCCAGCGTTGTCTAAAA 395
 QY 181 GCGTTGTTTATGAGCTCAAAAGAAAAATAAGACTTCTCTACTGTGGATTACACACCAA 240
 Db IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII

396 GTATTTTATGAGTTCAGANGAATAAAGACTTCGCTGCTGTGGATTGAANAACCAA 455
 QY 241 GAAGACAGAAACCACTGTCTCTCGAAATAATAGTAAAGCTTTT 281
 Db IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII

456 GCGGCGAGAGATGAGCTATACTGCAAAAACAATTAACCTTTT 496

RESULT 15

AAF90510
 ID AAF90510 standard; cDNA; 1172 BP.

XX AAF90510;

XX 22-AUG-2001 (first entry)

XX Caenorhabditis elegans core 1 beta3-galactosyl transferase cDNA.

XX Core 1 beta3-galactosyl transferase; nematode; O-glycosylation;
 KW galactosylation; glycosulfopeptide; Tn syndrome; IgA nephropathy;
 KW diagnosis; therapy; ss.

XX Caenorhabditis elegans.

OS Key Location/Qualifiers
 XX CDS 1..1170

FT /*tag= a
 FT /note= "a polynucleotide comprising this
 FT coding sequence is also claimed in Claim
 FT 1(A)"

XX WO200144478-A2.

XX 21-JUN-2001.

XX 14-DEC-2000; 2000WO-US33945.

XX 15-DEC-1999; 99US-0461321.

PR 15-DEC-1999; 99US-0464035.

XX (UOVOK-) UNIV OKLAHOMA.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 23:09:36 ; Search time 41.5774 Seconds
(without alignments)
3606.885 Million cell updates/sec

Title: US-10-029-359A-3
Perfect score: 489
Sequence: 1 catctaaagactgatgaa.....gatatgtactgagcaagaa 489

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters:, 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*

2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*

3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*

4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*

5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq:*

6: /cgn2_6/ptodata/1/ina/backfiles.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| C 1 | 45.8 | 9.4 | 7218 | 1 | US-08-232-463-14 |
| 2 | 41.6 | 8.5 | 1827 | 4 | US-09-134-001C-1914 |
| 3 | 36 | 7.4 | 668 | 4 | US-09-605-785-367 |
| 4 | 36 | 7.4 | 668 | 4 | US-09-439-313-367 |
| 5 | 36 | 7.4 | 668 | 4 | US-09-062-451-293 |
| 6 | 36 | 7.4 | 668 | 4 | US-09-352-616A-367 |
| C 7 | 35 | 7.2 | 1569 | 3 | US-08-821-984-9 |
| C 8 | 35 | 7.2 | 1569 | 4 | US-09-328-749-9 |
| C 9 | 35 | 7.2 | 6895 | 1 | US-08-570-311-9 |
| C 10 | 35 | 7.2 | 6895 | 2 | US-08-353-485-9 |
| C 11 | 34.8 | 7.1 | 4967 | 4 | US-09-177-325-12 |
| 12 | 34.8 | 7.1 | 4967 | 4 | US-09-411-812A-12 |
| 13 | 34.8 | 7.1 | 4967 | 4 | US-09-590-113-12 |
| C 14 | 34.2 | 7.0 | 272 | 4 | US-09-221-017B-522 |
| 15 | 33.2 | 6.8 | 1175 | 4 | US-09-378-088A-77 |
| 16 | 33.2 | 6.8 | 1410 | 4 | US-09-378-088A-81 |
| C 17 | 33 | 6.7 | 80595 | 4 | US-09-078-294-3 |
| C 18 | 32.6 | 6.7 | 782 | 4 | US-09-221-017B-522 |
| 19 | 32.6 | 6.7 | 7685 | 4 | US-09-221-017B-1092 |
| C 20 | 32.4 | 6.6 | 99500 | 4 | US-09-798-096-10 |
| 21 | 32.2 | 6.6 | 424 | 4 | US-09-308-003-9 |
| 22 | 32.2 | 6.6 | 972 | 4 | US-09-134-001C-952 |
| 23 | 32.2 | 6.6 | 1071 | 4 | US-09-308-003-10 |
| 24 | 32.2 | 6.6 | 2685 | 4 | US-09-221-017B-616 |
| 25 | 32.2 | 6.6 | 10754 | 2 | US-08-966-958-1 |
| 26 | 32.2 | 6.6 | 10754 | 2 | US-09-215-817-1 |
| 27 | 32.2 | 6.6 | 10754 | 4 | US-09-342-353-1 |

| | | | | | | |
|------|------|-----|-------|---|---------------------|--------------------|
| 28 | 32 | 6.5 | 2223 | 1 | US-08-257-073-4 | Sequence 4, Appli |
| 29 | 31.8 | 6.5 | 2760 | 4 | US-09-198-484-1 | Sequence 1, Appli |
| 30 | 31.4 | 6.4 | 5053 | 4 | US-08-961-527-187 | Sequence 187, App |
| 31 | 31 | 6.3 | 1092 | 4 | US-09-221-017B-136 | Sequence 136, App |
| 32 | 31 | 6.3 | 1386 | 4 | US-09-221-017B-986 | Sequence 986, App |
| 33 | 31 | 6.3 | 4325 | 2 | US-08-888-497-21 | Sequence 21, Appli |
| 34 | 31 | 6.3 | 4325 | 4 | US-09-362-230-21 | Sequence 21, Appli |
| 35 | 31 | 6.3 | 4325 | 5 | PCT-US94-07926-21 | Sequence 21, Appli |
| C 36 | 31 | 6.3 | 6173 | 4 | US-09-221-017B-898 | Sequence 898, App |
| 37 | 30.8 | 6.3 | 1665 | 4 | US-09-134-001C-1565 | Sequence 1565, Ap |
| C 38 | 30.8 | 6.3 | 1972 | 1 | US-08-463-048-1 | Sequence 1, Appli |
| C 39 | 30.8 | 6.3 | 1972 | 1 | US-08-463-229-1 | Sequence 1, Appli |
| C 40 | 30.8 | 6.3 | 1972 | 2 | US-08-302-891-1 | Sequence 1, Appli |
| 41 | 30.8 | 6.3 | 20165 | 4 | US-09-609-816-7 | Sequence 7, Appli |
| 42 | 30.8 | 6.3 | 51259 | 3 | US-08-781-891-209 | Sequence 209, App |
| 43 | 30.6 | 6.3 | 3910 | 4 | US-09-302-620B-91 | Sequence 91, Appli |
| C 44 | 30.4 | 6.2 | 624 | 4 | US-09-318-978-9 | Sequence 9, Appli |
| C 45 | 30.2 | 6.2 | 18073 | 4 | US-09-078-294-12 | Sequence 12, Appli |

ALIGNMENTS

RESULT 1

US-08-232-463-14/c

; Sequence 14, Application US/08232463

; Patent No. 5670367

; GENERAL INFORMATION:

; APPLICANT: DORNER, F.

; APPLICANT: SCHEIFLINGER, F.

; APPLICANT: FALKNER, F. G.

; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

; NUMBER OF SEQUENCES: 52

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 1800 Diagonal Road, Suite 500

; CITY: Alexandria

; STATE: VA

; COUNTRY: USA

; ZIP: 22313-0299

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/232,463

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/935,313

; FILING DATE:

; APPLICATION NUMBER: EP 91 114 300.6

; FILING DATE: 26-AUG-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: BENT, Stephen A.

; REGISTRATION NUMBER: 29,768

; REFERENCE/DOCKET NUMBER: 30472/114 IMMU

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703)836-9300

; TELEFAX: (703)683-4109

; TELEX: 899149

; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 7218 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; CLONE: pTZ9pt-fls

US-08-232-463-14

Query Match 9.4%; Score 45.8; DB 1; Length 7218;

Best Local Similarity 7.9%; Pred. No. 0.0011;
Matches 35; Conservative 212; Mismatches 194; Indels 0; Gaps 0;

Qy 46 AAATACACAGACACAGAGTGTCTACACAACTCTATCAAGAGATCAAAATCTTTGCTGG 105
Db 1507 AAAAAACGCGATGTAGGCATCATCTAATACCTCTATCAAGATCTTAAAGAGATAG 1448

Qy 106 ATTATGACAGACCTCAAAATCTAGAAAAAAGATCAGACGCATCAGAGATACATGGGCC 165
Db 1447 AAGAAATTTGGTACRR 1388

Qy 166 CAGGTTCCATAAAGCGTTGTTATGAGCTCAAAAGAAATAAAGACTTCTCTACTGTG 225
Db 1387 RRR 1328

Qy 226 GGATTACACCAAGAGACAGACAACTGCTCTGCGAAATAGTTAAAGCTTTCTA 285
Db 1337 RRR 1268

Qy 286 TATGCTCATGACCATTTATCTGGAATACATGATGGTTTCATCAAGACAGATGATGATA 345
Db 1267 RRR 1208

Qy 346 TGTATATATACATGCGACAACTTGAATGGCTTCTCAAACTATAACCTGATGAA 405
Db 1207 RRR 1148

Qy 406 TCCACTTACTTTGGAAAGATTTAAGCACTGCAGAAACAGGACTACATGCGAGGA 465
Db 1147 RRR 1088

Qy 466 CGAGGATGATGACTGAGCAAA 486
Db 1087 RRR 1067

RESULT 2
US-09-134-001C-1914
; Sequence 1914, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1914
; LENGTH: 1827
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1914

Query Match 8.5%; Score 41.6; DB 4; Length 1827;
Best Local Similarity 51.3%; Pred. No. 0.01;
Matches 121; Conservative 0; Mismatches 114; Indels 1; Gaps 1;

Qy 189 TATGAGCTCAAAAGAAATAAGACTTCTCTACTGTGGATTACACACCAAGAGACAG 248
Db 1569 TTTTAGTAAATGAAGAACAATATCTCTCTGTTATTGATTAGAAAAATGAAA 1628

Qy 249 AAACCAACTGCTCGGAAATAGTTAAAGCTTTTCTATATGCTCATGACCACTGCGA 308
Db 1629 TTATGAATGTGATGGAAACAGCTACAGGTATAGTAAACGGGAATATAATGCGTG 1688

Qy 309 ATACATGATGGTTGATGAAGACAGATGATGATATATATATATATATGATGACAA 368
Db 1689 GTTGTAGGTACGTAATACAAATCATGATAAGTATTATTTTCTA-CACATTTATCAG 1747

Qy 369 CTTGAAATGGCTTCTCAAAACTATAACCCCTGATGAATCCACTTACTTTGGAAAA 424
Db 1748 ATGAAAAGCCTCTGGTGCAAAATGCTAAATTCATAAATGAAAAATATATAAGNA 1803

RESULT 3
US-09-605-785-367
; Sequence 367, Application US/09605785
; Patent No. 6321716
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C16
; CURRENT APPLICATION NUMBER: US/09/605,785
; CURRENT FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 835
; SOFTWARE: fastseq for Windows Version 3.0
; SEQ ID NO 367
; LENGTH: 668
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-605-785-367

Query Match 7.4%; Score 36; DB 4; Length 668;
Best Local Similarity 60.0%; Pred. No. 0.29;
Matches 60; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 25 ATTGCAATGCTAGTCATCATATAATACAGACACAGGTGTCAGTCACTGACAACTCTATCA 84
Db 297 ATTACATATCTTAGGAATTCAAAATAACATTCACAGCTTTCACCAACTAGTTATATTA 356

Qy 85 AAGATGAAATCTTTTCTGCTGATTATGACAGGACCTCAAA 124
Db 357 AAGGAGAAACTCATTTTATGCCATGTATTGAATCAAA 396

RESULT 4
US-09-439-313-367
; Sequence 367, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang, Yuqui
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael.
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.427C9


```

; Patent NO. 0203778
; GENERAL INFORMATION:
;
; APPLICANT: Tureci, Ozlem
; APPLICANT: Sahin, Ugur
; APPLICANT: Pireundschuh, Michael
; TITLE OF INVENTION: Isolated Nu
; TITLE OF INVENTION: Thereof

```

```

MOLECULE TYPE: DNA (genomic)
FEATURE:
; APPLICANT: Pireundschun, Michael
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Encode SCP Proteins, A
; TITLE OF INVENTION: Thereof

```

FILE REFERENCE: LUD 5525.1 CIP
CURRENT APPLICATION NUMBER: US/09/411.812A
CURRENT FILING DATE: 1999-10-01
PRIOR APPLICATION NUMBER: US 09/177,325
PRIOR FILING DATE: 1998-10-22
NUMBER OF SEQ ID NOS: 14
SEQ ID NO 12
LENGTH: 4967
TYPE: DNA
ORGANISM: Homo sapiens
US-09-411-812A-12

Query Match 7.1%; Score 34.8; DB 4; Length 4967;
Best Local Similarity 48.5%; Pred. No. 1.2;
Matches 96; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 232 CACACCAAGAGACAGACAGAACCAACTGCTCGGAAATAGTTAAAGCTTTTCTATATGCT 291
DB 4555 CAGAGAACTGATGCAGTATTCATGCTCATGAAGAAATGCTAATGCTGAAATCTAGT 4614
QY 292 CATGACCACTTCTCGGAATACATGATGGTTTCATGAAGCAGATGATGATATATGATA 351
DB 4615 TTTTATCACCATACTTTATCTAATTTATTTCTCTGTATATACTGAGGAATAGATA 4674
QY 352 TATATCATTGGACAACTTGAATGGCTTCTCACAACACTATAACCCCTGATGAATCCACT 411
DB 4675 GTCTACAAAGAGAAAAATATACATGTCACCGAAGCAAGTGTACCCCTTTATAGGAACCT 4734
QY 412 TACTTTGGGAAAAGATTT 429
DB 4735 CAAATTAATAAAAAATGT 4752

RESULT 13
US-09-590-113-12
Sequence 12, Application US/09590113
Patent No. 6306389
GENERAL INFORMATION:
APPLICANT: Tureci, Ozlem
APPLICANT: Sakin, Ugur
APPLICANT: Pfeunderschuh, Michael
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Encode SCP Proteins, And Us
FILE REFERENCE: LUD 5525
CURRENT APPLICATION NUMBER: US/09/590.113
CURRENT FILING DATE: 2000-06-08
PRIOR APPLICATION NUMBER: US 09/177,325
PRIOR FILING DATE: 1998-10-22
NUMBER OF SEQ ID NOS: 15
SEQ ID NO 12
LENGTH: 4967
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
US-09-590-113-12

Query Match 7.1%; Score 34.8; DB 4; Length 4967;
Best Local Similarity 48.5%; Pred. No. 1.2;
Matches 96; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 232 CACACCAAGAGACAGACAGAACCAACTGCTCGGAAATAGTTAAAGCTTTTCTATATGCT 291
DB 4555 CAGAGAACTGATGCAGTATTCATGCTCATGAAGAAATGCTAATGCTGAAATCTAGT 4614
QY 292 CATGACCACTTCTCGGAATACATGATGGTTTCATGAAGCAGATGATGATATATGATA 351
DB 4615 TTTTATCACCATACTTTATCTAATTTATTTCTCTGTATATACTGAGGAATAGATA 4674
QY 352 TATATCATTGGACAACTTGAATGGCTTCTCACAACACTATAACCCCTGATGAATCCACT 411
DB 4675 GTCTACAAAGAGAAAAATATACATGTCACCGAAGCAAGTGTACCCCTTTATAGGAACCT 4734
QY 412 TACTTTGGGAAAAGATTT 429

DB 4735 CAAATTAATAAAAAATGT 4752
RESULT 14
US-09-221-017B-617/c
Sequence 617, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221.017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Monroy, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 617:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...272
US-09-221-017B-617

Query Match 7.0%; Score 34.2; DB 4; Length 272;
Best Local Similarity 58.3%; Pred. No. 0.7;
Matches 60; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 165 CCAGGGTTGCAATAAAGCGTTGTTTATGAGCTCAAAAGAAATAAAGACTTCTCTACTGT 224
DB 147 CCATGGGAGGATATTTATGCTGTTTTTAATACTCAATATACAAATACTCCCTACTAT 88
QY 225 GGGATTACACACCAAGAGACAGAACCAACTGTCTCTGGAAA 267

Db 87 ACAATCAATTAAACAGCAAAATACACACCAATTCGGTGC AAA 45

RESULT 15
US-09-378-088A-77
; Sequence 77, Application US/09378088A
; Patent No. 6372480
; GENERAL INFORMATION:
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Schnepf, H. Ernest
; APPLICANT: Knuth, Mark
; APPLICANT: Pollard, Michael R.
; APPLICANT: Cardineau, Guy A.
; APPLICANT: Schwab, George E.
; APPLICANT: Michaels, Tracy E.
; APPLICANT: Finstad Lee, Stacy
; APPLICANT: Burneister, Paula
; APPLICANT: Dojillo, Joanna
; TITLE OF INVENTION: Pesticidal Proteins
; FILE REFERENCE: MA703C2
; CURRENT APPLICATION NUMBER: US/09/378,088A
; CURRENT FILING DATE: 1999-08-20
; PRIOR APPLICATION NUMBER: US 08/633,993
; PRIOR FILING DATE: 1996-04-19
; PRIOR APPLICATION NUMBER: US 08/844,188
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 77
; LENGTH: 1175
; TYPE: DNA
; ORGANISM: Bacillus thuringiensis
US-09-378-088A-77

Query Match 6.8%; Score 33.2; DB 4; Length 1175;
Best Local Similarity 51.5%; Pred. No. 2.1;
Matches 101; Conservative 0; Mismatches 93; Indels 2; Gaps 1;
QY 2 ATCTAAAAGACTGATGAAGTTGATTGCAAAATGCTAGTCATCATAAATACCAGAACACAG 61
Db 744 ATTTCAAATTAATGTAGATTCCAGGAATGAAGTTTGAGGTACCGAAGTAGGAGGTAC 803
QY 62 GTGTCACCTGACAACTCTATCAAAAGATCAAAATTTCTTTGCTGGATTATGACAGGACCTC 121
Db 804 AGAGAGAAATAAAACACAAATTAATGA--AGAATTAAAGTTGAATATAGCACTGACACC 861
QY 122 AAAATCTAGAAAAAAGATCAGCCGATCAGAGATACATGGCCCCCAGGGTTGCAATAAAG 181
Db 862 AAAATAATGAAAAAATATCAAGAACACTCAGAGATAGATAATCCCACTAATCAACAATG 921
QY 182 CGTTGTTTATGAGCTC 197
Db 922 AATCTATAGGATTTC 937

Search completed: June 28, 2003, 00:24:01
Job time : 43.5774 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 23:29:41 ; Search time 126.594 seconds
(without alignments)
5734.045 Million cell updates/sec

Title: US-10-029-359A-3
Perfect score: 489
Sequence: 1 catctaaagactgatgaa.....gatattgactgagcaagaa 489

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1055720 seqs, 742224136 residues

Total number of hits satisfying chosen parameters: 2111440

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_applications_NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| SUMMARIES | | | | Description | |
|------------|-------|-------------|--------------|---------------------|-------------------|
| Result No. | Score | Query Match | Length DB ID | | |
| 1 | 489 | 100.0 | 489 9 | US-10-029-359A-3 | Sequence 3, Appli |
| C 2 | 458.6 | 93.8 | 555 9 | US-10-198-846-13278 | Sequence 13278, A |
| C 3 | 415.6 | 85.0 | 879 9 | US-10-198-846-6422 | Sequence 6422, Ap |
| 4 | 307 | 62.8 | 1092 9 | US-10-235-056-12 | Sequence 12, Appl |
| 5 | 307 | 62.8 | 1794 9 | US-10-235-056-2 | Sequence 2, Appli |
| 6 | 305.4 | 62.5 | 1560 9 | US-10-029-359A-5 | Sequence 5, Appli |
| 7 | 300.6 | 61.5 | 1092 9 | US-10-235-056-13 | Sequence 13, Appl |
| 8 | 300.6 | 61.5 | 1440 9 | US-10-235-056-4 | Sequence 4, Appli |
| 9 | 292.6 | 59.8 | 1092 9 | US-10-235-056-14 | Sequence 14, Appl |
| 10 | 292.6 | 59.8 | 1469 9 | US-10-235-056-6 | Sequence 6, Appli |
| C 11 | 278 | 56.9 | 1975 10 | US-09-864-761-15628 | Sequence 15628, A |
| C 12 | 277.4 | 56.7 | 1770 10 | US-09-864-761-32135 | Sequence 32135, A |
| 13 | 105 | 21.5 | 1170 9 | US-10-235-056-15 | Sequence 15, Appl |
| 14 | 105 | 21.5 | 1172 9 | US-10-235-056-8 | Sequence 8, Appli |
| 15 | 92.2 | 18.9 | 1167 9 | US-10-235-056-18 | Sequence 18, Appl |
| 16 | 74.8 | 15.3 | 1104 9 | US-10-235-056-10 | Sequence 10, Appl |
| 17 | 46.2 | 9.4 | 299 9 | US-10-029-359A-1 | Sequence 1, Appli |
| 18 | 39.2 | 8.0 | 409 10 | US-09-783-590-8460 | Sequence 8460, Ap |
| 19 | 38.6 | 7.9 | 640681 10 | US-09-790-988-1 | Sequence 1, Appli |

| | | | | | | |
|------|------|-----|---------|----|---------------------|-------------------|
| C 20 | 37 | 7.6 | 161 | 10 | US-09-969-373-1319 | Sequence 1319, Ap |
| C 21 | 36.6 | 7.5 | 652 | 10 | US-09-746-801A-38 | Sequence 38, Appl |
| C 22 | 36.2 | 7.4 | 126512 | 10 | US-09-804-474A-3 | Sequence 3, Appli |
| 23 | 36 | 7.4 | 668 | 9 | US-09-924-400-293 | Sequence 293, App |
| 24 | 36 | 7.4 | 668 | 9 | US-10-012-896-367 | Sequence 367, App |
| 25 | 36 | 7.4 | 668 | 9 | US-09-895-793-367 | Sequence 367, App |
| 26 | 36 | 7.4 | 668 | 9 | US-09-895-814-367 | Sequence 367, App |
| 27 | 36 | 7.4 | 668 | 9 | US-10-010-940-367 | Sequence 367, App |
| 28 | 36 | 7.4 | 668 | 10 | US-09-759-143-367 | Sequence 367, App |
| 29 | 36 | 7.4 | 668 | 10 | US-09-780-669-367 | Sequence 367, App |
| 30 | 36 | 7.4 | 668 | 10 | US-09-810-936-293 | Sequence 293, App |
| 31 | 36 | 7.4 | 668 | 10 | US-09-822-827-367 | Sequence 367, App |
| 32 | 36 | 7.4 | 668 | 10 | US-09-429-755-293 | Sequence 293, App |
| C 33 | 35.4 | 7.2 | 640681 | 10 | US-09-790-988-1 | Sequence 9, Appli |
| C 34 | 35 | 7.2 | 1569 | 10 | US-09-849-566-9 | Sequence 9, Appli |
| C 35 | 35 | 7.2 | 1569 | 10 | US-09-907-859-9 | Sequence 9, Appli |
| 36 | 34.8 | 7.1 | 468 | 9 | US-09-918-995-13121 | Sequence 13121, A |
| 37 | 34.8 | 7.1 | 515 | 9 | US-09-796-692-7170 | Sequence 7170, Ap |
| 38 | 34.8 | 7.1 | 515 | 9 | US-10-040-862-7170 | Sequence 7170, Ap |
| 39 | 34.8 | 7.1 | 5456 | 10 | US-09-834-975-971 | Sequence 971, App |
| C 40 | 34.8 | 7.1 | 1503841 | 9 | US-09-946-807-1 | Sequence 1, Appli |
| C 41 | 34.8 | 7.1 | 1503841 | 10 | US-09-795-668-1 | Sequence 1, Appli |
| C 42 | 34.8 | 7.1 | 1503841 | 10 | US-09-795-668-1 | Sequence 1, Appli |
| C 43 | 34.6 | 7.1 | 533 | 9 | US-09-796-692-3419 | Sequence 3419, Ap |
| C 44 | 34.6 | 7.1 | 533 | 9 | US-10-040-862-3419 | Sequence 3419, Ap |
| C 45 | 34.6 | 7.1 | 605 | 10 | US-09-811-284-113 | Sequence 113, App |

ALIGNMENTS

RESULT 1
US-10-029-359A-3
; Sequence 3, Application US/10029359A
; Publication No. US20030100056A1
; GENERAL INFORMATION:
; APPLICANT: Attersand, Anelli
; TITLE OF INVENTION: Protein Cluster II.
; FILE REFERENCE: 10806-156
; CURRENT APPLICATION NUMBER: US/10/029,359A
; CURRENT FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 489
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (55)..(489)
; OTHER INFORMATION:
US-10-029-359A-3

| | | | | | |
|---------------------------|-----|---|-----|--------------------------------------|--|
| Query Match | | | | 100.0%; Score 489; DB 9; Length 489; | |
| Best Local Similarity | | | | 100.0%; Pred. No. 8.2e-124; | |
| Matches 489; Conservative | | | | 0; Mismatches 0; Indels 0; Gaps 0; | |
| QY | 1 | CATCTAAAGAGCTGATGAAGTTGATTGCAAAATCTCTATCAATATACCAACACACA | 60 | | |
| Db | 1 | CATCTAAAGAGCTGATGAAGTTGATTGCAAAATCTCTATCAATATACCAACACACA | 60 | | |
| QY | 61 | GGTGTCACTGACAACTCTATCAAAAGATGAAAATCTTTGCTGGATTATGACAGACCT | 120 | | |
| Db | 61 | GGTGTCACTGACAACTCTATCAAAAGATGAAAATCTTTGCTGGATTATGACAGACCT | 120 | | |
| QY | 121 | CAAAATCTAGAAAAAAGATGACAGCATCAGATACATATGCGCCCGGGTTTCAATAA | 180 | | |
| Db | 121 | CAAAATCTAGAAAAAAGATGACAGCATCAGATACATATGCGCCCGGGTTTCAATAA | 180 | | |
| QY | 181 | GCCTGTTTATGAGCTCAAAAGAAAATAAGACTTCTTCTACTGTGGATTACACACAAA | 240 | | |
| Db | 181 | GCCTGTTTATGAGCTCAAAAGAAAATAAGACTTCTTCTACTGTGGATTACACACAAA | 240 | | |

Qy 241 GAAGACAGAAACCAACTGCTCTCGAAATAAGTTAAAGCTTTTCTATATGCTCATGACCAT 300
Db 241 GAAGACAGAAACCAACTGCTCTCGAAATAAGTTAAAGCTTTTCTATATGCTCATGACCAT 300
Qy 301 TATCTGGNATACATGGATTGTTTCATGAAACGACATGATATATATGATATATATACACA 360
Db 301 TATCTGGNATACATGGATTGTTTCATGAAACGACATGATATATATGATATATATACACA 360
Qy 361 TTGGACAACCTTGAATGGCTTCTCACAACTATATACCTGATGATATACCTTACTTTGGG 420
Db 361 TTGGACAACCTTGAATGGCTTCTCACAACTATATACCTGATGATATACCTTACTTTGGG 420
Qy 421 AAAAGATTAAAGCACTGCAGAAAACAGGACTACATGACTGGAGGAGGAGGATATGTA 480
Db 421 AAAAGATTAAAGCACTGCAGAAAACAGGACTACATGACTGGAGGAGGAGGATATGTA 480
Qy 481 AGCAAGAA 489
Db 481 AGCAAGAA 489

RESULT 2

US-10-198-846-13278/c

; Sequence 13278, Application US/10198846

; Publication No. US20030099974A1

; GENERAL INFORMATION:

; APPLICANT: Lillie, James

; APPLICANT: Xu, Yongyao

; APPLICANT: Wang, Youzhen

; APPLICANT: Steinmann, Kathleen

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS

; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND

; TITLE OF INVENTION: THERAPY OF BREAST CANCER

; FILE REFERENCE: MRI-049

; CURRENT APPLICATION NUMBER: US/10/198,846

; CURRENT FILING DATE: 2002-07-18

; PRIOR APPLICATION NUMBER: 60/306,220

; PRIOR FILING DATE: 2001-07-18

; NUMBER OF SEQ ID NOS: 14084

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13278

; LENGTH: 555

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc.feature

; LOCATION: 553..554, 555

; OTHER INFORMATION: n = A,T,C or G

US-10-198-846-13278

Query Match 93.8%; Score 458.6; DB 9; Length 555;
Best Local Similarity 99.0%; Pred. No. 1.8e-115;
Matches 472; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
Qy 1 CATCTAAAAGACTGATGAAGTTGATGCAAACTGCTAGTCATCATATAATACCAACACA 60
Db 477 CATCTAAAAGACTGATGAAGTTGATGCAAACTGCTAGTCATCATATAATACCAACACA 418
Qy 61 GGTCTCACTGACAACCTCTATCAAAAGATGAAATTTCTTCTGGATATGACAGGACCT 120
Db 417 GGTCTCACTGACAACCTCTATCAAAAGATGAAATTTCTTCTGGATATGACAGGACCT 358
Qy 121 CAAATCTAG-AAAAAAGATCAGACGCTCAGAGATACATGGCCCGAGGTTGCAATAA 179
Db 357 CAAATCTAGAAAAAAGATCAGACGCTCAGAGATACATGGCCCGAGGTTGCAATAA 298
Qy 180 AGCGTTGTTATGAGCTCAAAAGAAAAATAAGACTTCTTCTACTGTGGATTACACACAA 239
Db 297 AGCGTTGTTATGAGCTCAAAAGAAAAATAAGACTTCTTCTACTGTGGATTACACACAA 238
Qy 240 AGAAGACAGAACCAACTGCTCTGAAAAATAGTTAAAGCTTTTCTATATGCTCATGACCA 299
Db 237 AGAAGACAGAACCAACTGCTCTGAAAAATAGTTAAAGCTTTTCTATATGCTCATGACCA 178

Qy 300 TTATCTGAAATACATGATTGGTTTCATGAAAGACAGATGATGATATATGATATATACAC 359
Db 177 TTATCTGAAATACATGATTGGTTTCATGAAAGACAGATGATGATATATGATATATACAC 118
Qy 360 ATTGGACAACCTGAAATGGCTTCTCACAACTATATAACCTGATGAATCCACTTACTTTGG 419
Db 117 ATTGGACAACCTGAAATGGCTTCTCACAACTATATAACCTGATGAATCCACTTACTTTGG 58
Qy 420 GAAAGATTTAAAGCACTGCAGAAAACAGGACTACATGACTGGAGGAGGAGGATATGT 476
Db 57 GAAAGATTTAAAGCACTGCAGAAAACAGGACTACATGACTGGAGGAGGAGGATATGT 1

RESULT 3

US-10-198-846-6422/c

; Sequence 6422, Application US/10198846

; Publication No. US20030099974A1

; GENERAL INFORMATION:

; APPLICANT: Lillie, James

; APPLICANT: Xu, Yongyao

; APPLICANT: Wang, Youzhen

; APPLICANT: Steinmann, Kathleen

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS

; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND

; TITLE OF INVENTION: THERAPY OF BREAST CANCER

; FILE REFERENCE: MRI-049

; CURRENT APPLICATION NUMBER: US/10/198,846

; CURRENT FILING DATE: 2002-07-18

; PRIOR APPLICATION NUMBER: 60/306,220

; PRIOR FILING DATE: 2001-07-18

; NUMBER OF SEQ ID NOS: 14084

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 6422

; LENGTH: 879

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc.feature

; LOCATION: 41..71, 120, 187, 212, 226, 230, 239, 264, 275, 277, 287,

; LOCATION: 358, 373, 390, 395, 420, 436, 492, 495, 497, 519, 530, 556,

; LOCATION: 594, 616, 634, 652, 662, 664, 671, 673, 675, 677, 682, 688,

; LOCATION: 689, 693, 695, 698, 700, 704, 708, 712, 715, 728, 731

; OTHER INFORMATION: n = A,T,C or G

; FEATURE:

; NAME/KEY: misc.feature

; LOCATION: 732, 734, 737, 742, 750, 752, 754, 755, 768, 772, 773, 785,

; LOCATION: 788, 794, 798, 800, 805, 808, 810, 826, 828, 831, 839, 844,

; LOCATION: 851, 853, 859, 869, 874, 879

; OTHER INFORMATION: n = A,T,C or G

US-10-198-846-6422

Query Match 85.0%; Score 415.6; DB 9; Length 879;
Best Local Similarity 93.8%; Pred. No. 1.2e-103;
Matches 453; Conservative 0; Mismatches 25; Indels 5; Gaps 3;
Qy 1 CATCTAAAAGACTGATG-AGTTTGGATTGCAAAATGCTAGTCATCA--TAAATACCAAC 57
Db 509 CATCTAAAAGAGATNATNTAGTTGATTTGCAATGCTAGTCATCAATTAATTCACAAAC 450
Qy 58 ACAGGTGCTCACTGACAAACTCTATCAAAAGATGAAATTTCTTCTGCTGGATTATGACAGGA 117
Db 449 ACAGGTGCTCACTGACAAACTCTATCAAAATGAAATTTCTTCTGCTGGATTATGACAGGN 390
Qy 118 CCTCAAAATCT--AGAAAAAAGATCAGACGCTCAGAGATACATGCGCCCGAGGTTGCA 175
Db 389 CCTCAAAATCTTAGAANAANAAGATCAGACGATACATGCGCCCGAGGTTGCA 330
Qy 176 ATAAAGCGTTGTTTATGAGCTCAAAAGAAAAATAAGACTTCTTCTACTGTGGATTACACA 235
Db 329 ATAAAGCGTTGTTTATGAGCTCAAAAGAAAAATAAGACTTCTTCTACTGTGGNTACACA 270
Qy 236 CCAAAGAGACAGAAACCAACTGTCTCTGAAAAATAGTTAAAGCTTTTCTATATGCTCATG 295

Db 269 CCAANAAGACAGAAACCAACTGCTCTCGANAATAGTTANAGCNITTCATATGCTCTG 210
QY 296 ACCATTATCTGGAATACATGGATTGTTTCATGAAGCAGATGATGATATATATATA 355
Db 209 ACCATTATCTGGAATACATGGATTGTTTCATGAAGCAGATGATGATATATATATA 150
QY 356 TCACATTTGGACAACTTGAATGGCTTCTCACAACTATATAACCTGATGAATCCACTTACT 415
Db 149 TCACATTTGGACAACTTGAATGGCTTCTCNCAACTATATAACCTGATGAATCCACTTACT 90
QY 416 TTGGGAAAAGATTAAAGCACTGCAGAAAACAGACTACATGACTGGAGGAGCAGGATATG 475
Db 89 TTGGGAAAAGATTAAAGCNCCTGCAGAAAACAGACTACATGACTGGAGNAGCAGGATATG 30
QY 476 TAC 478
Db 29 TAC 27

RESULT 4

US-10-235-056-12
; Sequence 12, Application US/10235056
; Publication No. US2003005915A1
; GENERAL INFORMATION:
; APPLICANT: Canfield, William M.
; APPLICANT: Cummings, Richard D.
; APPLICANT: Ju, Tongzhong
; TITLE OF INVENTION: CORE 1 3-GALACTOSYL TRANSFERASES
; FILE REFERENCE: 5820.630
; CURRENT APPLICATION NUMBER: US/10/235,056
; PRIOR FILING DATE: 2002-09-04
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 1092
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-235-056-12

Query Match 62.8%; Score 307; DB 9; Length 1092;
Best Local Similarity 78.3%; Pred. No. 6.1e-74;
Matches 383; Conservative 0; Mismatches 100; Indels 6; Gaps 1;

QY 1 CATCTAAAAGACGATGAGCTTGATTGCAAACTGCTAGTATCATATAATACCAGACACA 60
Db 172 CATCTAGAAGACAAATGAACCTTCAATGCAGATTCTAGCCAAATTAAGATGAGAACACA 231
QY 61 GGTGTCACTGACAACTCTATCAAAAGATGAAAATCTTTGCTGGATTATGACAGACCT 120
Db 232 GACATTGCTGAAAACCTCTATCAGAAAAGTTAGAATCTTCTGCTGGTTATGACCGCCCT 291
QY 121 CAAATCTAGAAAAAAGATCAGACGATCAGATACATATGCGGCCCGAGGTTGCAATAAA 180
Db 292 CAAACCTAGAAAAAAGCCCAACGCTCAAGACTACTTGGGCCCGAGCCTTTGTAACAAA 351
QY 181 GCGTTGTTTATGAGCTCAAAAGAAAATAAGACTTCTCTACTGTGGGATTACACACAAA 240
Db 352 GTGTTGTTTATGAGTTTCAAGAAAATAAAGACTTCCCTGCTGGGACTGAAAACCAA 411
QY 241 GAAGACAGAACCAACTGCTCTGAAAATAGTTAAAGCTTTTCHVATATGCTCATGACCAT 300
Db 412 GAAGCAGAGATCAACTATATCTGAAAACAATTAAGCTTTTTCAGTATGTTTCAATGAACAT 471
QY 301 TATCTGGAATACATGGAATGGTTTCATGAAGCAGATGATATATATATATATACACA 360
Db 472 TATTAGAGATGCTGATTGGTTTTCAGAAAGATGATGA-----CAGCTATGTCATA 525
QY 361 TTGACAACTTGAATGGCTTCTCAAACTATAACCCCTGATGAATCCACTTACTTTGGG 420
Db 526 CTAGACAAATTGAGGTGGCTCTTCAAAATACGCCCTGAAGACCCATTTACTTTGGG 585

QY 421 AAAAGATTAAAGCACTGTCAGAAAACAGACTACATGCTGGAGGAGCAGGATATGACTG 480
Db 586 AGAAGATTAAAGCCTTATGTAAGCAGCGCTACATGAGTGGAGGAGCAGGATATGACTA 645
QY 481 AGCAAAAGAA 489
Db 646 AGCAAAAGAA 654

RESULT 5

US-10-235-056-2
; Sequence 2, Application US/10235056
; Publication No. US2003005915A1
; GENERAL INFORMATION:
; APPLICANT: Canfield, William M.
; APPLICANT: Cummings, Richard D.
; APPLICANT: Ju, Tongzhong
; TITLE OF INVENTION: CORE 1 3-GALACTOSYL TRANSFERASES
; FILE REFERENCE: 5820.630
; CURRENT APPLICATION NUMBER: US/10/235,056
; CURRENT FILING DATE: 2002-09-04
; PRIOR FILING DATE: US 09/461,321
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1794
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-235-056-2

Query Match 62.8%; Score 307; DB 9; Length 1794;
Best Local Similarity 78.3%; Pred. No. 7.7e-74;
Matches 383; Conservative 0; Mismatches 100; Indels 6; Gaps 1;

QY 1 CATCTAAAAGACGATGAGCTTGATTGCAAACTGCTAGTATCATATAATACCAGACACA 60
Db 234 CATCTAGAAGACAAATGAACCTTCAATGCAGATTCTAGCCAAATTAAGATGAGAACACA 293
QY 61 GGTGTCACTGACAACTCTATCAAAAGATGAAAATCTTTGCTGGATTATGACAGACCT 120
Db 294 GACATTGCTGAAAACCTCTATCAGAAAAGTTAGAATCTTTCCTGCTGGTTATGACCGCCCT 353
QY 121 CAAATCTAGAAAAAAGATCAGACGATCAGATACATGCGGCCCGAGGTTGCAATAAA 180
Db 354 CAAACCTAGAAAAAAGCCCAACGCTCAAGACTACTTGGGCCCGAGCCTTTGTAACAAA 413
QY 181 GCGTTGTTTATGAGCTCAAAAGAAAATAAGACTTCTCTACTGTGGGATTACACACAAA 240
Db 414 GTGTTGTTTATGAGTTTCAAGAAAATAAAGACTTCCCTGCTGGGACTGAAAACCAA 473
QY 241 GAAGACAGAACCAACTGCTCTGAAAATAGTTAAAGCTTTTCTATATGCTCATGACCAT 300
Db 474 GAAGCAGAGATCAACTATATCTGAAAACAATTAAGACTTTTTCAGTATGTTTCAATGAACAT 533
QY 301 TATCTGGAATACATGGAATGGTTTCATCAAGCAGATGATATATATATATATACACA 360
Db 534 TATTAGAGATGCTGATTGGTTTTCAGAAAGATGATGA-----CAGCTATGTCATA 587
QY 361 TTGACAACTTGAATGGCTTCTCAAACTATAAACCCTGATGAATCCACTTACTTTGGG 420
Db 588 CTAGACAAATTGAGGTGGCTTCTTTCAAATACAGCCCTGAAGACCCATTTACTTTGGG 647
QY 421 AAAAGATTAAAGCACTGCAGAAAACAGGACTACATGACTGGAGGAGCAGGATATGACTG 480
Db 648 AGAAGATTAAAGCCTTATGTAAGCAGCGCTACATGAGTGGAGGAGCAGGATATGACTA 707
QY 481 AGCAAAAGAA 489
Db 708 AGCAAAAGAA 716

RESULT 6

US-10-029-359A-5
; Sequence 5, Application US/10029359A
; Publication No. US20030100056A1
; GENERAL INFORMATION:
; APPLICANT: Attersand, Anneli
; TITLE OF INVENTION: Protein Cluster II
; FILE REFERENCE: 10806-156
; CURRENT APPLICATION NUMBER: US/10/029,359A
; PRIORITY FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 1560
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)..(934)
; OTHER INFORMATION:
US-10-029-359A-5

Query Match 62.5%; Score 305.4; DB 9; Length 1560;
Best Local Similarity 78.1%; Pred. No. 2e-73;
Matches 382; Conservative 0; Mismatches 101; Indels 6; Gaps 1;

```
QY 1 CATCTAAAGAGACTGATGAAGTTGATTCGAAATGCTAGTCATCATATAAATACAGAACACA 60
    ||||| || || ||||| || || ||||| || || ||||| || || ||||| || || |||||
DB 17 CATCTAGAGGCAATGAACCTCAATCGAGATTCAGCCACATATAAGATGAGAACACA 76
    ||||| || || ||||| || || ||||| || || ||||| || || ||||| || || |||||
QY 61 GGTGTCACGACAACTCTATCAAAAGATGAAAATCTTTGCTGGATTATGACAGACCT 120
    || || || || || || || || || || || || || || || || || || || || || ||
DB 77 GACATGCTGAAACCTCTATCAGAAAGCTTAGAATTCCTTGTGGTTATGACGCGCCT 136
    || || || || || || || || || || || || || || || || || || || || || ||
QY 121 CAAATCTAGAAAAAGATCAGACGCATCAGAGATACATGGGCCCGGGTTGCAATAAA 180
    ||||| ||||| ||||| || || || || || || || || || || || || || || || ||
DB 137 CAAACCTAGAAAAAGGCCCAACACGCTCAAGCTACTTGGGCCCGCGTTGTAAACAA 196
    || || || || || || || || || || || || || || || || || || || || || ||
QY 181 GCGTTGTTATGAGCTCAAAAGATAAAGACTTCTCTACTGTGGGATTACACACCAA 240
    ||||| ||||| || || || || || || || || || || || || || || || || || ||
DB 197 GTGTTGTTATGAGTTCAGAGAAATAAAGACTTCCCTGCTGTGGGACTGAAACCAA 256
    || || || || || || || || || || || || || || || || || || || || || ||
QY 241 GAAGACAGAAACCAACTCTCTCGGAAATAGTAAAGCTTTTCTATATGCTCATGACCAT 300
    ||||| ||||| || || || || || || || || || || || || || || || || || ||
DB 257 GAAGCAGAGATCAACTACTCTGGAACAAATTAAGCTTTTTCAGTATGTTTCATGAACAT 316
    || || || || || || || || || || || || || || || || || || || || || ||
QY 301 TATCTGGAATACATGGATTGGTTTCATGAAGCAGATGATATATATATATATACACA 360
    || || || || || || || || || || || || || || || || || || || || || ||
DB 317 TATTTACAAGATGCTGATTGGTTTGTGAAGCAGATGATGA-----CACGTATGTCATA 370
    || || || || || || || || || || || || || || || || || || || || || ||
QY 361 TTGGACAACTTGAATGGCTTCTCAAACTATTAACCTGATGAATCCACTTACTTTGGG 420
    ||||| ||||| || || || || || || || || || || || || || || || || || ||
DB 371 CTAGACAATTTGAGTGGCTCTTTTCAAAATACGACCTGAAGAACCCTTACTTTGGG 430
    || || || || || || || || || || || || || || || || || || || || || ||
QY 421 AAAAGATTAAAGCATGCGAGAAACAGGACTACATGCTGAGGAGCAGGATATGTTACTG 480
    ||||| ||||| || || || || || || || || || || || || || || || || || ||
DB 431 AGAAGATTAAAGCCTTATGTAAGAGCGGCTACATGAGTGGAGGAGCAGGATATGTTACTA 490
    || || || || || || || || || || || || || || || || || || || || || ||
QY 481 AGCAAGAA 489
    ||||| |||||
DB 491 AGCAAGAA 499
```

RESULT 7

US-10-235-056-13
; Sequence 13, Application US/10235056
; Publication No. US20030059915A1
; GENERAL INFORMATION:
; APPLICANT: Canfield, William M.
; APPLICANT: Cummings, Richard D.
; APPLICANT: Ju, Tongzhong
; TITLE OF INVENTION: CORE 1 3-GALACTOSYL TRANSFERASES
; FILE REFERENCE: 5820.630

; CURRENT APPLICATION NUMBER: US/10/235,056
; CURRENT FILING DATE: 2002-09-04
; PRIORITY FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 1092
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-10-235-056-13

Query Match 61.5%; Score 300.6; DB 9; Length 1092;
Best Local Similarity 77.5%; Pred. No. 3.4e-72;
Matches 379; Conservative 0; Mismatches 104; Indels 6; Gaps 1;

```
QY 1 CATCTAAAAAGACTGATGAAGTTGATTCGAAATGCTAGTCATCATATAAATACAGAACACA 60
    || || || || || || || || || || || || || || || || || || || || || ||
DB 172 CACCTCAAAGGACAGATGGACTTCAATGCAGATTCAGCCACATATAAGATGAGAACACA 231
    ||||| ||||| ||||| || || || || || || || || || || || || || || || ||
QY 61 GGTGTCACGACAACTCTATCAAAAGATGAAAATCTTTGCTGGATTATGACAGACCT 120
    || || || || || || || || || || || || || || || || || || || || || ||
DB 232 GACGTTGCTGAGAACCTCTATCAGAAAGTTAAAGTTCCTTGTGGTTATGACAAGCCCT 291
    ||||| ||||| ||||| || || || || || || || || || || || || || || || ||
QY 121 CAAATCTAGAAAAAGATCAGACGCATCAGAGATACATGGGCCCGGGTTGCAATAAA 180
    ||||| ||||| ||||| || || || || || || || || || || || || || || || ||
DB 292 CAAATCTAGAAAAAGGCCCAACACGCTCAAGCTACATGGGCCCGCGTTGTAATAAA 351
    ||||| ||||| ||||| || || || || || || || || || || || || || || || ||
QY 181 GCGTTGTTATGAGCTCAAAAGATAAAGACTTCTCTACTGTGGGATTATACACACCAA 240
    ||||| ||||| || || || || || || || || || || || || || || || || || ||
DB 352 GTGTTATTTATGAGTTCAGAAAGATAAAGACTTCCCTACTGTGGGCTGGAACCCAAA 411
    ||||| ||||| || || || || || || || || || || || || || || || || || ||
QY 241 GAAGACAGAAACCAACTCTCTCGGAAATAGTAAAGCTTTTCTATATGCTCATGACCAT 300
    ||||| ||||| ||||| || || || || || || || || || || || || || || || ||
DB 412 GAAGCAGAGACCAACTCTACTTGGAAAAACAATTAAGCGTTTTCAGTACGTACATGACCAT 471
    || || || || || || || || || || || || || || || || || || || || || ||
QY 301 TATCTGGAATACATGGATTGGTTTCATGAAGCAGATGATATATATATATATATACACA 360
    || || || || || || || || || || || || || || || || || || || || || ||
DB 472 TACTTAGAAGATGCTGACTGGTTTATGAAGCAGACGATGA-----CACCTATGTCATA 525
    || || || || || || || || || || || || || || || || || || || || || ||
QY 361 TTGGACAACCTTGAATGGCTTCTCAAACTATTAACCTGATGAATCCACTTACTTTGGG 420
    ||||| ||||| || || || || || || || || || || || || || || || || || ||
DB 526 CTGGACAATCTGAGATGGCTTCTATCAAGATATAACCTGTAACAGCCCATTTACTTTGGG 585
    ||||| ||||| || || || || || || || || || || || || || || || || || ||
QY 421 AAAAGATTAAAGCATGCGAGAAACAGGACTACATGCTGAGGAGCAGGATATGTTACTG 480
    ||||| ||||| || || || || || || || || || || || || || || || || || ||
DB 586 AGAAGATTAAAGCCCTATGTGAAGCAGGGATACATGAGTGGAGGAGCAGGATATGTCCTA 645
    || || || || || || || || || || || || || || || || || || || || || ||
QY 481 AGCAAGAA 489
    ||||| |||||
DB 646 AGCAAGAA 654
```

RESULT 8

US-10-235-056-4
; Sequence 4, Application US/10235056
; Publication No. US20030059915A1
; GENERAL INFORMATION:
; APPLICANT: Canfield, William M.
; APPLICANT: Cummings, Richard D.
; APPLICANT: Ju, Tongzhong
; TITLE OF INVENTION: CORE 1 3-GALACTOSYL TRANSFERASES
; FILE REFERENCE: 5820.630
; CURRENT APPLICATION NUMBER: US/10/235,056
; CURRENT FILING DATE: 2002-09-04
; PRIORITY FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1440
; TYPE: DNA
; ORGANISM: Rattus norvegicus

US-10-235-056-4

| | | | | | | | |
|-----------------------|--------|--------------|----------|------------|------|--------|-------|
| Query Match | 61.5%; | Score | 300.6; | DB | 9; | Length | 1440; |
| Best Local Similarity | 77.5%; | Pred. No. | 3.9e-72; | | | | |
| Matches | 379; | Conservative | 0; | Mismatches | 104; | Indels | 6; |
| Gaps | 1; | | | | | | |

| | | | |
|----|-----|---|-----|
| QY | 1 | CATCTAAAAGAGACTGATGAAGTTGATTGCAAAATGCTAGTCATCATATAAATACAGAACACA | 60 |
| Db | 325 | CACCTCAAGGACAGATGGACTTCAATCAGATTCTAGCCAAACATAAAGATGAGAACACA | 384 |
| QY | 61 | GGTGCTACTGCACAAACTCTATCAAAAGATGAAATTCCTTCTGCTGGATTATCACAGGACCT | 120 |
| Db | 385 | GACGTTGCTGAGAACCTCTATCAGAAAGTTAAAGTTCCTTTGGGTTATGACAAGCCCT | 444 |
| QY | 121 | CAAAATCTAGAAAAAAGATCAGACGCATCAGAGATACATGGGCCCGAGGTTGCAATAAA | 180 |
| Db | 445 | CAAAATCTAGAAAAAAGGCCCAACACGCTCAAAGCTACATGGGCCCGAGGTTGTAATAA | 504 |
| QY | 181 | CGTTGTTTATGAGCTCAAAAGAAAAATAAGACATCTCTTATATGCTCATGACCCAT | 300 |
| Db | 505 | GTGTTATTATGAGTTTCAGAAAAATAAGACATCTCTTATATGCTCATGACCCAT | 624 |
| QY | 241 | GAAGACGAACACCACTGTCTCTGGAAAAATAGTTTAAAGCTTTTCTATATGCTCATGACCCAT | 360 |
| Db | 565 | GAAGCAGAGAGCAACTGTACTGGAACCAATTAAGCGTTTCAGTACGTACATGACCAT | 624 |
| QY | 301 | TATCTGGAATACATGGATTGGTTTCATGAAGCAGATGATGATATGATATATATACACA | 420 |
| Db | 625 | TACTTGAAGATGCTGACTGGTTTATGAAAGCAGACGATGA-----CACCTATGTCATA | 678 |
| QY | 361 | TTGACAACTTGAATGGCTTCTCACAAACTATAACCTGATGAATCCACTTACTTTGGG | 420 |
| Db | 679 | CTGCACAACTGAGATGGCTTCTATCAAGATATAACCTGAACAGCCCATTTACTTTGGG | 738 |
| QY | 421 | AAAAGATTTAAGCACTCGACAAACAGGACTACATGACTCGAGGAGCAGGATATGTACTG | 480 |
| Db | 739 | AGAAGATTTAAGCCCTATGTGAAGCAGGATACATGATGAGTGAGGAGCAGGATATGTCTA | 798 |
| QY | 481 | AGCAAAAGAA | 489 |
| Db | 799 | AGCAAGAA | 807 |

| | |
|-------------------------------|----------------------------------|
| RESULT | 9 |
| US-10-235-056-14 | |
| ; Sequence | 14, Application US/10235056 |
| ; Publication | No. US20030059915A1 |
| ; GENERAL INFORMATION: | |
| ; APPLICANT: | Canfield, William M. |
| ; APPLICANT: | Cummings, Richard D. |
| ; APPLICANT: | Ju, Tongzhong |
| ; TITLE OF INVENTION: | CORE 1 3-GALACTOSYL TRANSFERASES |
| ; FILE REFERENCE: | 5820.630 |
| ; CURRENT APPLICATION NUMBER: | US/10/235,056 |
| ; CURRENT FILING DATE: | 2002-09-04 |
| ; PRIOR APPLICATION NUMBER: | US 09/461,321 |
| ; PRIOR FILING DATE: | 1999-12-15 |
| ; NUMBER OF SEQ ID NOS: | 25 |
| ; SOFTWARE: | PatentIn version 3.1 |
| ; SEQ ID NO | 14 |
| ; LENGTH: | 1092 |
| ; TYPE: | DNA |
| ; ORGANISM: | Mus musculus |
| US-10-235-056-14 | |

| | | |
|-----|--|-----|
| 531 | GTGCTATTATTCAGTTCGGAAGAAATCAAGACTTCCTACTGTGGGATTGAAACCAA | 590 |
| Db | | |
| 241 | GAAGACAGAAACCAACTGTCTCGGAAAATAGTTAAAGCTTTTCTATATGCTCATGACCAT | 300 |
| Qy | | |
| 591 | GAAGGCAGAGACAACTATTGTGAAAACAATTTAAAGCTTTCAGTATGTATGATGACCAT | 650 |
| Db | | |
| 301 | TATCTGGAGATACATGATTGTTTCATGAAACGACATGATATATCTATATATATATACAA | 360 |
| Qy | | |
| 651 | TATTTAGAGATGCTGACTGGTTTATGAAACGATGACGA-----CACATACGTCATT | 704 |
| Db | | |
| 361 | TTGGACAACCTTGAATGGCTTCTCAACACTATTAACCTTGATGAATCCACTTACTTTGGG | 420 |
| Qy | | |
| 705 | GTGGACAACCTGAGATGGCTTCTCAAAAGTATAACCTTGAACAACCCATTACTTTGGG | 764 |
| Db | | |
| 421 | AAAAGATTTAAGCACTGCAGAAAACAGAGCTACATGACTGGAGGACGAGATATGTACTG | 480 |
| Qy | | |
| 765 | CGAAGATTTTAAAGCCCTATGTGAACAGGATACATGAGCGGAGGAGCGGGCTATGTCCTA | 824 |
| Db | | |
| 481 | AGCAAGAA | 489 |
| Qy | | |
| 825 | AGCAAGGA | 833 |
| Db | | |

RESULT 11

```

US-09-864-761-15628/c
; Sequence 15628, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomics-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29

```

```

; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.11
; SEQ ID NO 15628
; LENGTH: 1975
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005332.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =
US-09-864-761-15628

```

| Query Match | 56.9%; | Score 278; | DB 10; | Length 1975; |
|-----------------------|-----------------|--|-----------|--------------|
| Best Local Similarity | 78.4%; | Pred. No. 6.7e-66; | | |
| Matches 348; | Conservative 0; | Mismatches 90; | Indels 6; | Gaps 1 |
| Qy | 46 | AAATACCAGAACACAGCTGTGCACCTGACACAACTCTATCAAAAAGATGAAATTTCTTTGCTGG | 105 | |
| Db | 938 | ATAGATCAGAAACACAGACATTCCTGAAACCTCTATCAGAAAGTTAGAAATTTCTTTGCTGG | 879 | |
| Qy | 106 | ATTATGACAGGACCTCAAAATCTAGAAAAAAGATCAGACGCATCAGAGATACATGCGCC | 165 | |
| Db | 878 | GTTATGACCGGCCCTCAAAACCTAGAGAAAAAGGCCAACACAGTCAAAAGCTACTTTGGGCC | 819 | |
| Qy | 166 | CAGGGTTGCAATAAAGCGTGTGTTTATGAGCTCAAAAGAAAAATAAAGACTTCTCTACTGTG | 225 | |
| Db | 818 | CAGCGTTGTAAACAAGTGTGTTTATGAGTTCAGAGAGAAAAATAAAGACTTCCCTGCTGTG | 759 | |
| Qy | 226 | GGATTTACACACCACAAAGACAGACAAACCAACTGCTCTGGAAAAATAGTTAAAGCTTTTCTA | 285 | |
| Db | 758 | GGACTGAAACACCAAGAAAGGCAGAGATCAACTATACTGGAAAACAATTTAAAGCTTTTTCAG | 699 | |
| Qy | 286 | TATGCTCATGACCATTATCTGGATATACATGGATTGGTTCATGAAGACAGATGATGATATA | 345 | |
| Db | 698 | TATGTTCTATGAACATATATTTAGAGATGCTGATTTGGTTTTTTTGAAGACAGATGATGA | 643 | |
| Qy | 346 | TGTATATATATCATTTGGACAACTTGAATTGGCTTCTCACAAACTATAAACCCTGATGAA | 405 | |
| Db | 642 | --CACGTATGTCTACTATGACAAATTTGAGGTGGCTTCTTTCAAATACGACCCCTGAAGAA | 585 | |
| Qy | 406 | TCCACTTACTTTGGGAAAGATTTTAAACACTTGCACAAACAGGACTACATGACTGGAGGA | 465 | |
| Db | 584 | CCCATTTACTTTGGGAGAAAGATTTAAGCCCTATGTAAAGCAGGCGTACATGAGTGGAGGA | 525 | |
| Qy | 466 | GCAGGATATGTACTGAGCAAGAA | 489 | |
| Db | 524 | GCAGGATATGTACTTAAGCAAGAA | 501 | |

RESULT 12

```

US-09-864-761-32135/c
; Sequence 32135, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27

```


;; PRIOR FILING DATE: 1999-12-15
;; NUMBER OF SEQ ID NOS: 25
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 8
;; LENGTH: 1172
;; TYPE: DNA
;; ORGANISM: Caenorhabditis elegans
US-10-235-056-8

Query Match 21.58; Score 105; DB 9; Length 1172;

Best Local Similarity 55.28; Pred. No. 9.4e-19;
Matches 22; Conservative 0; Mismatches 180; Indels 6; Gaps 1;

QY 68 CTGACAAACTCTATCAAAAGATGAAATCTTTGCTGGATTATGACAGGACCTCAAAATC 127
DB 290 CCGAGGAAGTTGCGAAGAGTTCCGCTCTTCTGTTGGATTCTCACCGGAAACAGATC 349
QY 128 TAGAAAAAGATGACAGACGATCAGAGATACATGAGGCCCCAGGGTTGCAATAAAGCGTTGT 187
DB 350 ACGATAAAGCGGCGGAAACAGCTCAAGCCACCTGGGCGCAAGCGGTGTAATAAGTACGTAT 409
QY 188 TTATGACCTCAAGAGAAATTAAGACTTCTACTCTGTGGGATTACACACCAAAAGACACA 247
DB 410 TCATGTCATCAGAGAGAGATGCGAAGACTCCAGCCATCAACCTAAATGTCTCTGAAGGCA 469
QY 248 GAAACCAACTGCTCTGGAATAATAGTTAAAGCTTTTCTATATGCTCATGACCATTTATCTGG 307
DB 470 GAGATTACTTGTGGGCAAAACAAAGAGGAGCATTCATCAATTAATGACCATCACCTGA 529
QY 308 AATACATGGATTGGTTTCATGAAGCAGATGATGATATATATATATATATATATATATGACAT 367
DB 530 ACGACTACGACTGGTCTCTGAAAGCCGACGACGA- - - - -TACCTATGTGTGATGAGAA 583
QY 368 ACTTGAATGCTCTCTCACAACTATAACCTTGATGATCCACTTACTTTGGGAAAGAT 427
DB 584 ATCTCCGGTTTCATGCTATTGGCTATTCCACCGGATGAGCCAAATTCACCTTTGGATGCAAGT 643
QY 428 TTAAGCACTGCAGAAAAACAGACTACATGATGCTGGAGGAGGAGGATGATGATGAG 482
DB* 644 TTAAGCCATTACACAGGAGGATATCATATGTTGGAGCTGGTGGAGCTGGATATGCTCAG 698

RESULT 15

US-10-235-056-18
;; Sequence 18, Application US/10235056
;; Publication No. US20030059915A1
;; GENERAL INFORMATION:
;; APPLICANT: Cummings, William M.
;; APPLICANT: Ju, Tongzhong
;; TITLE OF INVENTION: CORE 1 3-GALACTOSYL TRANSFERASES
;; FILE REFERENCE: 5820.630
;; CURRENT APPLICATION NUMBER: US/10/235.056
;; CURRENT FILING DATE: 2002-09-04
;; PRIOR APPLICATION NUMBER: US 09/461,321
;; PRIOR FILING DATE: 1999-12-15
;; NUMBER OF SEQ ID NOS: 25
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 18
;; LENGTH: 1167
;; TYPE: DNA
;; ORGANISM: Drosophila melanogaster
US-10-235-056-18

Query Match 18.98; Score 92.2; DB 9; Length 1167;

Best Local Similarity 52.4%; Pred. No. 3e-15;
Matches 230; Conservative 0; Mismatches 203; Indels 6; Gaps 1;

QY 50 ACCAGAACACAGGTGTCTACTGACAACTATCAAAAGATGAAATTTCTTTGCTGGATTA 109
DB 257 ACGAGACTCGACCATTTGGCGGAGGACTGTACAGCGAGGTGCGTCTCTGCTGGATCA 316
QY 110 TGACAGGACCTCAAAATCTAGAAAAAAGATCAGACCCATCAGAGATACATGGGCCAGG 169

DB 317 TGACCAATCCGAGCAACCATCAGAAGAAGCGGCCACGTCGAAAGCGCACCTGGGGCAAGC 376
QY 170 GTTGCATTAAGCGTTGTTTATGAGCTCAAAAGAAATTAAGACTTCTCTACTCTGGAT 229
DB 377 GTTGCACAAAGCTGATCTTTATGAGCTCCGCCAAGGACGACGAGCTGGAGCGAGTGGCTC 436
QY 230 TACACACCAAAAGAGACAGAAACCAACTGCTCTGGAAATAGTTAAAGCTTTTCTATATG 289
DB 437 TGGCCGTAGCGAGGTCGCAACAACCTATGGGGCAGAGCAAGGAGGCCCTACAATACA 496
QY 290 CTCATGACCATTTATCTGGAATACATGATGTTGTTTCATGAAAGCAGATGATGATATGTA 349
DB 497 TCTACGAGCATCACATCAACGACGCGGACTGTTCTCTGAAGGCTGACGATGA- - - - -CA 550
QY 350 TATATATACATTTGGACAACCTTGAATGGCTTCTCACAACTATATACCTTGATGAATCCA 409
DB 551 CATACAGATAGTGGAGAACATCGGATACATGCTGTATCCGTACAGTCCGGGAAACTCCAG 610
QY 410 CTTACTTTGGGAAAAGATTTAAGCACTTGCAGAAAAACAGGACTACATGACTGGAGGAGCAG 469
DB 611 TCTACTTCGCTCAAGTTCAAGCCGTACGTGAACAAGCTACATGTCCGGCGGTGCCG 670
QY 470 GATATGTTACTGAGCAAGA 488
DB 671 GCTACGTTCTCAGCCGGA 689

Search completed: June 28, 2003, 00:27:37

Job time : 129.594 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 23:34:12 ; Search time 1231.81 Seconds
(without alignments)
6429.241 Million cell updates/sec

Title: US-10-029-359A-3

Perfect score: 489

Sequence:

1 catctaaaagactgatgaa.....gatatgactgagcaagaag 489

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:**
1: em_estba:**
2: em_esthum:**
3: em_estlin:**
4: em_estmu:**
5: em_estov:**
6: em_estpl:**
7: em_estro:**
8: em_htc:**
9: gb_est1:**
10: gb_est2:**
11: gb_htc:**
12: gb_est3:**
13: gb_est4:**
14: gb_est5:**
15: em_estfun:**
16: em_estom:**
17: gb_gss:**
18: em_gss_hum:**
19: em_gss_inv:**
20: em_gss_pln:**
21: em_gss_vrt:**
22: em_gss_fun:**
23: em_gss_nam:**
24: em_gss_mus:**
25: em_gss_other:**
26: em_gss_pro:**
27: em_gss_rod:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-------------|--------------------|
| 1 | 311.4 | 63.7 | 481 | 12 BF443807 | BF443807 261354 MA |
| 2 | 264.2 | 54.0 | 696 | 13 BI102329 | BI102329 602885661 |
| 3 | 260 | 53.2 | 637 | 10 BE336820 | BE336820 ba97a03.y |
| 4 | 221.2 | 45.2 | 614 | 9 AJ449471 | AJ449471 AJ449471 |
| 5 | 177.6 | 36.3 | 642 | 9 AJ447748 | AJ447748 AJ447748 |
| 6 | 167.4 | 34.2 | 528 | 17 AQ230580 | AQ230580 HS_2033_A |

```

7 163 33.3 505 14 N47633
8 144.6 29.6 1051 17 CNS018X7
c 9 143.4 29.3 594 12 BF471960
c 10 141.2 28.9 480 17 BI5559
11 131.8 27.0 606 12 BG077752
12 126.2 25.8 1088 17 CNS058Y8
13 124.2 25.4 915 14 BQ736824
14 111 22.7 582 13 BI508514
15 111 22.7 646 13 BI509237
16 108.6 22.2 618 17 AZ456919
17 98.4 20.1 316 12 BE993641
18 96.4 19.7 497 9 AL712318
19 92. 18.8 501 9 AI546714
20 91.6 18.7 458 9 AI964359
21 89.4 18.3 651 13 BJ114209
22 87 17.8 677 10 AW773412
23 85.2 17.4 462 9 AI236007
24 81 16.6 471 9 AI059600
25 79.8 16.3 474 9 AI546718
26 73.4 15.0 257 10 BB602602
27 72 14.7 518 13 BI450723
28 70.6 14.4 434 12 BF938171
29 70.4 14.4 557 14 BM714438
30 70.2 14.4 328 12 BF158206
31 67.6 13.8 544 13 BI502334
32 66.4 13.6 563 13 BJ040104
33 64.8 13.3 506 13 BM533050
34 64.2 13.1 399 13 BM289607
c 35 64 13.1 689 9 AI816690
36 59 12.1 1101 17 CNS0082F
37 56.6 11.6 614 10 AW419812
38 55.4 11.3 352 14 R03773
39 55 11.2 141 14 T10488
40 55 11.2 570 9 AL698324
41 53.8 11.0 559 13 BI240060
42 52.4 10.7 941 17 CNS006VT
43 52.2 10.7 737 9 AI387726
44 52.2 10.7 797 17 CNS03JD9
c 45 51 10.4 1101 17 CNS014PC

```

ALIGNMENTS

```

RESULT 1
BF443807
LOCUS 261354 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
DEFINITION BF443807
ACCESSION BF443807
VERSION BF443807.1 GI:11503836
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 481)
AUTHORS Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
and Keele,J.W.
TITLE Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT

```


✓


```

VERSION BF471960.1 GI:11541143
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE Bonaïdo, M.F., Lennon, G. and Soares, M.B.
NORMALIZATION Normalization and subtraction: two approaches to facilitate gene
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd., Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
cDNA Library Preparation: M.B. Soares Lab Clone Distribution:
Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
should be noted that Bento Soares is generating a small number of
additional specialized non-redundant arrays of BMAP cDNAs whose
availability will be considered under appropriate and limited
collaborative arrangements. The following repetitive elements were
found in this cDNA sequence: 565-594, >GC-rich#Low_complexity
Seq primer: M13 Reverse.

FEATURES             Location/Qualifiers
     .source          1..594
     .organism         "Mus musculus"
     .strain           "C57BL/6J"
     .db_xref          "taxon:10090"
     .clone            "UI-M-BH3-awu-c-04-0-UI"
     .clone_lib        "NIH_BMAP_M_S4"
     .dev_stage        "27-32 days"
     .lab_host         "DH10B (Life Technologies)"
     .notes            "Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The
NIH_BMAP_M_S4 library is a subtracted library of a series,
ultimately derived from a mixture of individually tagged
normalized libraries from ten regions of the mouse brain
(cerebellum, brain stems, olfactory bulbs, hypothalamus,
cortex, amygdala, basal ganglia, pineal gland, striatum,
hippocampus) after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: NIH_BMAP_M_S4,
NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library
(NIH_BMAP_M_S4) was constructed as follows: PCR amplified
cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
was used as a driver in a hybridization with a pool of
the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
libraries in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library)
was purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
NIH_BMAP_M_S4 library. This procedure has been previously
described (Bonaïdo, Lennon and Soares, Genome Research
6:791-806, 1996)".

BASE COUNT 158 a 136 c 148 g 152 t
ORIGIN

Query Match 29.3%; Score 143.4; DB 12; Length 594;
Best Local Similarity 77.3%; Pred. No. 2.9e-25;
Matches 174; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 1 CATCTAAAGACGTGATGATGCTGCAATCTAGTCTATCATATAATACGACACACA 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 235 CACCTCAAGAGGACAGATGATGCAATCTAGTCTAGTCTATCATATAATACGACACATA 176

RESULT 10
LOCUS B15559/c
DEFINITION 345H11.TP CIT978SK1 Homo sapiens genomic clone A-345H11, DNA
sequence.
ACCESSION B15559
VERSION B15559.1 GI:21233308
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Adams, M.D., Kelley, J.M., Rounsley, S.R. and Venter, J.C.
JOURNAL Use of a BAC End Sequence Database for Sequence-Ready Map Building
COMMENT Unpublished (1997)
Other_GSSs: 345H11.TV
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: SP6
Class: BAC ends.

FEATURES             Location/Qualifiers
     .source          1..480
     .organism         "Homo sapiens"
     .db_xref          "taxon:9606"
     .clone            "A-345H11"
     .clone_lib        "CIT978SK1"
     .sex              "Female"
     .cell_type        "Fibroblast"
     .note             "Vector: pBAC108L; Site 1: HindIII; Site 2: HindIII;
Caltech Human BAC Library A1"

BASE COUNT 144 a 90 c 75 g 171 t
ORIGIN

Query Match 28.9%; Score 141.2; DB 17; Length 480;
Best Local Similarity 78.0%; Pred. No. 9.9e-25;
Matches 170; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 46 AAATACCAGAACACAGGTGTCTACTGACAAAACCTCTATCAAAAGATGAAAATCTTCTGCTGG 105
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 218 ATAGATGAGAACACACAGACATTTGCTGAAACCTCTATCAGAAAGTTAGAAATCTTTGCTGG 159

QY 106 ATTATGACAGGACCTCAAAATCTAGAAAAAAGATCAGACGCATCAGAGATACATGGGCC 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 158 GTTATGACCGGCCCTCAAAACCTTAGAGAAAAAGCCCAACACGCTACTTGGGCC 99

QY 166 CAGGTTGCAATAAAGCGTTGTTTATGAGCTCAAAAGAAAATAAAGACTTCTCTACTGTG 225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 98 CAGCGTTGTAAACAAGTGTGTTTATGAGTTCAGAGAGAAAATAAAGACTTCTCTGCTGTG 39

QY 226 GGATTACACACCAGAACAGACAGAAACCAACTGTCTCTG 263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


QY 128 TAGAAAAAAGATCAGAGCGCATCAGAGATACATGGGCCCGCCAGGTTGCAATAAAGCGTTGT 187
Db 250 TGGAGACCAAGCGCGGCAGCTCAAAANCACTGGACGCGCCACTGCAACACATTTGGTTT 309
QY 188 TTATGAGCTCAAAAGAAAATAAAGACTTCTCTACTGTGGGATTTACACACCAAAAGACA 247
Db 310 TCATGAGCTCGTGGAGGACCCCAACTTCCCAACCGTGGGCTGGTACGAAAGAGGGGC 369
QY 248 GAAACCAACTCTCTCGGAAAATAGTTAAAGCTTTTCTATATGCTCATGACCATATCTGG 307
Db 370 GGGACCACTCTTCTGGAAANCACTCCGGGCTTCCACTAGCCCTACGAGCACCACTGCG 429
QY 308 AATACATGATGTTGTTCAAGAGAGATGATATATATATATATATATATATATATATATAT 367
Db 430 ACAGGCGGATTTGTTCTCAAGCGGAGCATGA-----CACCTATGTGATGAGACA 483
QY 368 ACTTGAATGCTTCTCAACAATATATACCTGATGATGATGATGATGATGATGATGATGATG 427
Db 484 ACTCCCGCTGGGCTCTCGCCCAACACACGCGCCGAGCCCATCTNCTTCGCGCGGAG-T 542
QY 428 TTAAGCACTGCAGAAAAACAGGACTACATGACTGAGGAGGAGGATATGCTACTGAGCAAAG 487
Db 543 TCAACCTCNCACCAAGAGGCTACATGTCGCGCGGCGGCTACGCTCTTGAGTAAG 602
QY 488 A 488
Db 603 A 603

RESULT 13
LOCUS B0736824 915 bp mRNA linear EST 16-JUL-2002
DEFINITION AGENCOURT 8115982 NICH D XCC Emb4 xenopus laevis cDNA clone
IMAGE:5571880 5', mRNA sequence.
ACCESSION B0736824
VERSION B0736824.1 GI:21875721
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 915)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Dr. Igor Dawid
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.lnl.gov/bbrp/image/image.html
Plate: LLAM12316 row: g column: 17
High quality sequence stop: 656.
Location/Qualifiers
1. .915
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:5571880"
/clone_lib="NICH D XCC Emb4"
/dev_stage="embryo, stage 31-32"
/lab_host="DH10B (phage-resistant)"
/note="Organ: whole embryo; Vector: pCMV-SPORT6; Site:1;
Noti; Site:2; Sali; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 2.1 kb. Constructed by Life
Technologies. Note: This is a Xenopus Gene Collection (XCC
library)."
262 a 217 c 188 g 248 t

FEATURES
source

Query Match 25.4%; Score 124.2; DB 14; Length 915;
Best Local Similarity 60.3%; Pred. No. 1.8e-20;
Matches 243; Conservative 0; Mismatches 153; Indels 7; Gaps 2;
QY 24 GATTGCAAACTAGTTCATCAATAATACCAAGAACAGGCTGCTCACTGACAACTCTATCA 83
Db 428 GTTTGAATGTCCCATTTTTCATCTGCCAGAAGATAATAGTGTGAGTGAGGAACCTCTCCA 487
QY 84 AAAGATGAAATCTTCTGCTGGATATGACAGGACCTCAAAATCTAGAAAAAAGATCAG 143
Db 488 AAAGGTGAGAGTGTGTTGTTGATCATGACTGGACCGCAACCTAAAGACCAAGCTAT 547
QY 144 ACGCATCAGAGATACATGGGCCCGGTTGCAATAAAGCGTTGTTATGAGCTCAAAAGA 203
Db 548 TCATGTGAGAAATTCCTGGACCGCTCACTGCAATGTAGCGTATTCATGAGCTCCATCAC 607
QY 204 AAATAAGACTTCTCTACTGTGGGATTACACACCAAGAACAGACAAACCACTGCTCTG 263
Db 608 CGATGAAGACTTCCCGCCATTTGCTGGGAACCTGGGGAAGGACAGACAACTTTACTG 667
QY 264 GAAATAGTAAAGCTTTTCTATATGCTCATGACCATATCTGGAATACA-TGGATTGGT 322
Db 668 GAAACAAATTCGTGCTTCCATTACGCCCATTAAGTACTACCTGAATGAGACAGAAATGGT 727
QY 323 TCATGAAAGCAGATGATGAT 382
Db 728 TCTTCAAGCAGATGATGA-----TACATATGAATTAATATGGAATTTACGATGGATGC 781
QY 383 TCACAACTATACCCCTGATGATCCACTTACTTTGGGAAAAG 425
Db 782 TTTTCAACTATACCGCTGACCGAGCTATTTTACTTTGGGAAAAG 824
RESULT 14
LOCUS B1508514 582 bp mRNA linear EST 08-APR-2002
DEFINITION BBI70009B21H02.5 Bee Brain Normalized/Subtracted Library, BBI7 Apis
mellifera cDNA clone BBI70009B21H02 5', mRNA sequence.
ACCESSION B1508514
VERSION B1508514.1 GI:15358888
KEYWORDS EST.
SOURCE honeybee.
ORGANISM Apis mellifera
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
Apidae; Apis.
1 (bases 1 to 582)
Whitfield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L.,
Pardinas,J., Robertson,H.M., Soares,B. and Robinson,G.E.
Annotated expressed sequence tags and cDNA microarrays for studies
of brain and behavior in the honey bee
Genome Res. 12 (4), 555-566 (2002)
21929762
Contact: Gene E. Robinson
Department of Entomology
University of Illinois
505 S. Goodwin Ave., Urbana, IL 61801, USA
Tel: 217 265 0309
Fax: 217 244 3499
Email: generobi@life.uiuc.edu
This research was funded by the University of Illinois Critical
Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
Award in Functional Genomics to G.E. Robinson and an NSF
Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
PCR Primers
FORWARD: TAATACGACTCACTATAGG
BACKWARD: ATTAACCTCACTAAAG
Plate: BBI70009B21 row: H column: 02
Seq primer: AGCGATAACAATTTACACAGGA
High quality sequence stop: 582.
Location/Qualifiers
1. .582
FEATURES
source

